

DOCKET NO. : CHIR-0160
(356.001)

PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In Re Application of:

Enzo Scalato, Vega Masignani, Rino
Rappuoli, Mariagrazia Pizza,, and Guido
Grandi

Serial No.: not assigned

Group Art Unit: not assigned

Filing Date: April 30, 1999

Examiner: not assigned

For: Neisserial Antigens

EXPRESS MAIL LABEL NO: EL066389665US

DATE OF DEPOSIT: April 30, 1999

Box ☒ Patent Application
☐ Provisional ☐ Design ☐ Sequence

Assistant Commissioner for Patents
Washington DC 20231

Sir:

PATENT APPLICATION TRANSMITTAL LETTER

Transmitted herewith for filing, please find

☒ A Utility Patent Application under 37 C.F.R. 1.53(b).

It is a continuing application, as follows:

☐ continuation ☐ divisional ☒ continuation-in-part of prior application number
PCT/IB98/01665 filed October 9, 1998.

☐ A Provisional Patent Application under 37 C.F.R. 1.53(c).

- ☐
- A Design Patent Application (submitted in duplicate).

Including the following:

- ☐
- Provisional Application Cover Sheet.

- ☒ New or Revised Specification, including pages 1 to 518 containing:

- ## Specification

- ☒
- Claims

- ☒
- Abstract

- ☐
- Substitute Specification, including Claims and Abstract.

- ☐ The present application is a continuation application of Application No. _____ filed _____. The present application includes the Specification of the parent application which has been revised in accordance with the amendments filed in the parent application. Since none of those amendments incorporate new matter into the parent application, the present revised Specification also does not include new matter.

- ☐ The present application is a continuation application of Application No. _____ filed _____, which in turn is a continuation-in-part of Application No. _____ filed _____. The present application includes the Specification of the parent application which has been revised in accordance with the amendments filed in the parent application. Although the amendments in the parent C-I-P application may have incorporated new matter, since those are the only revisions included in the present application, the present application includes no new matter in relation to the parent application.

- ☐ A copy of earlier application Serial No. _____ Filed _____, including Specification, Claims and Abstract (pages 1 - @@), to which no new matter has been added TOGETHER WITH a copy of the executed oath or declaration for such earlier application and all drawings and appendices. Such earlier application is hereby incorporated into the present application by reference.

- ☐ Please enter the following amendment to the Specification under the Cross-Reference

to Related Applications section (or create such a section): "This Application:

☐ is a continuation of ☐ is a divisional of ☐ claims benefit of U.S. provisional
Application Serial No. _____ filed _____

☐ Signed Statement attached deleting inventor(s) named in the prior application.

☐ A Preliminary Amendment.

☒ Twenty-seven (27) Sheets of ☐ Formal ☒ Informal Drawings.

☐ Petition to Accept Photographic Drawings.☐ Petition Fee

☒ An ☐ Executed ☒ Unexecuted Declaration or Oath and Power of Attorney.

☐ An Associate Power of Attorney.

☐ An ☐ Executed ☐ Copy of Executed Assignment of the Invention to _____

☐ A Recordation Form Cover Sheet.

☐ Recordation Fee - \$40.00.

☐ The prior application is assigned of record to _____

☒ Priority is claimed under 35 U.S.C. § 119 of Patent Application No. PCT/IB98/01665 filed October 9, 1998.

☐ A Certified Copy of each of the above applications for which priority is claimed:

☐ is enclosed.

☐ has been filed in prior application Serial No. _____ filed _____.

☐ An ☐ Executed or ☐ Copy of Executed Earlier Statement Claiming Small Entity

Status under 37 C.F.R. 1.9 and 1.27

☐ is enclosed.

☐ has been filed in prior application Serial No. _____ filed _____,
said status is still proper and desired in present case.

☐ Diskette Containing DNA/Amino Acid Sequence Information.

☐ Statement to Support Submission of DNA/Amino Acid Sequence Information.

☐ The computer readable form in this application _____, is identical with that filed in Application Serial Number _____, filed _____. In accordance with 37 CFR 1.821(e), please use the ☐ first-filed, ☐ last-filed or ☐ only computer readable form filed in that application as the computer readable form for the instant application. It is understood that the Patent and Trademark Office will make the necessary change in application number and filing date for the computer readable form that will be used for the instant application. A paper copy of the Sequence Listing is ☐ included in the originally-filed specification of the instant application, ☐ included in a separately filed preliminary amendment for incorporation into the specification.

☐ Information Disclosure Statement.

☐ Attached Form 1449.

☐ Copies of each of the references listed on the attached Form PTO-1449 are enclosed herewith.

☐ A copy of Petition for Extension of Time as filed in the prior case.

☐ Appended Material as follows: _____

☒ Return Receipt Postcard (should be specifically itemized).

☐ Other as follows: _____

00000160.01600000

FEE CALCULATION:

- ☐ Cancel in this application original claims _____ of the prior application before calculating the filing fee. (At least one original independent claim must be retained for filing purposes.)

				SMALL ENTITY		NOT SMALL ENTITY	
				RATE	FEE	RATE	FEE
PROVISIONAL APPLICATION				\$75.00	\$	\$150.00	\$
DESIGN APPLICATION				\$155.00	\$	\$310.00	\$
UTILITY APPLICATIONS BASE FEE				\$380.00	\$	\$760.00	\$ 760
UTILITY APPLICATION; ALL CLAIMS CALCULATED AFTER ENTRY OF ALL AMENDMENTS							
	No. Filed	No. Extra					
TOTAL CLAIMS	24 - 20 =	4	\$9 each	\$	\$18 each	\$ 72	
INDEP. CLAIMS	5 - 3 =	2	\$39 each	\$	\$78 each	\$ 156	
FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM			\$130	\$	\$260	\$ 260	
ADDITIONAL FILING FEE				\$		\$	
TOTAL FILING FEE DUE				\$		\$ 1248	

- ☒ A Check is enclosed in the amount of \$ 1,248.
- ☒ The Commissioner is authorized to charge payment of the following fees and to refund any overpayment associated with this communication or during the pendency of this application to deposit account 23-3050. This sheet is provided in duplicate.
- ☐ The foregoing amount due.
- ☒ Any additional filing fees required, including fees for the presentation of extra claims under 37 C.F.R. 1.16.
- ☒ Any additional patent application processing fees under 37 C.F.R. 1.17 or 1.20(d).
- ☐ The issue fee set in 37 C.F.R. 1.18 at the mailing of the Notice of Allowance.
- ☒ The Commissioner is hereby requested to grant an extension of time for the appropriate length of time, should one be necessary, in connection with this filing or

NEISSERIAL ANTIGENS

This application is a continuation-in-part of international patent application PCT/IB98/01665, filed October 9, 1998, from which priority is claimed under 35 U.S.C. § 119.

This invention relates to antigens from *Neisseria* bacteria.

5 BACKGROUND ART

Neisseria meningitidis and *Neisseria gonorrhoeae* are non-motile, gram negative diplococci that are pathogenic in humans. *N.meningitidis* colonises the pharynx and causes meningitis (and, occasionally, septicaemia in the absence of meningitis); *N.gonorrhoeae* colonises the genital tract and causes gonorrhea. Although colonising different areas of the body and causing completely different diseases, the two pathogens are closely related, although one feature that clearly differentiates meningococcus from gonococcus is the presence of a polysaccharide capsule that is present in all pathogenic meningococci.

N.gonorrhoeae caused approximately 800,000 cases per year during the period 1983-1990 in the United States alone (chapter by Meitzner & Cohen, "Vaccines Against Gonococcal Infection", In: *New Generation Vaccines*, 2nd edition, ed. Levine, Woodrow, Kaper, & Cobon, Marcel Dekker, New York, 1997, pp.817-842). The disease causes significant morbidity but limited mortality. Vaccination against *N.gonorrhoeae* would be highly desirable, but repeated attempts have failed. The main candidate antigens for this vaccine are surface-exposed proteins such as pili, porins, opacity-associated proteins (Opas) and other surface-exposed proteins such as the Lip, Laz, IgA1 protease and transferrin-binding proteins. The lipooligosaccharide (LOS) has also been suggested as vaccine (Meitzner & Cohen, *supra*).

N.meningitidis causes both endemic and epidemic disease. In the United States the attack rate is 0.6-1 per 100,000 persons per year, and it can be much greater during outbreaks (see Lieberman *et al.* (1996) Safety and Immunogenicity of a Serogroups A/C *Neisseria meningitidis* Oligosaccharide-Protein Conjugate Vaccine in Young Children. *JAMA* 275(19):1499-1503; Schuchat *et al* (1997) Bacterial Meningitis in the United States in 1995. *N Engl J Med* 337(14):970-976). In developing countries, endemic disease rates are much higher and during epidemics incidence rates can reach 500 cases per 100,000 persons per year. Mortality is extremely high, at

10-20% in the United States, and much higher in developing countries. Following the introduction of the conjugate vaccine against *Haemophilus influenzae*, *N. meningitidis* is the major cause of bacterial meningitis at all ages in the United States (Schuchat *et al* (1997) *supra*).

Based on the organism's capsular polysaccharide, 12 serogroups of *N.meningitidis* have been identified. Group A is the pathogen most often implicated in epidemic disease in sub-Saharan Africa. Serogroups B and C are responsible for the vast majority of cases in the United States and in most developed countries. Serogroups W135 and Y are responsible for the rest of the cases in the United States and developed countries. The meningococcal vaccine currently in use is a tetravalent polysaccharide vaccine composed of serogroups A, C, Y and W135. Although efficacious in adolescents and adults, it induces a poor immune response and short duration of protection, and cannot be used in infants [*eg.* Morbidity and Mortality weekly report, Vol.46, No. RR-5 (1997)]. This is because polysaccharides are T-cell independent antigens that induce a weak immune response that cannot be boosted by repeated immunization. Following the success of the vaccination against *H.influenzae*, conjugate vaccines against serogroups A and C have been developed and are at the final stage of clinical testing (Zollinger WD "New and Improved Vaccines Against Meningococcal Disease" in: *New Generation Vaccines, supra*, pp. 469-488; Lieberman *et al* (1996) *supra*; Costantino *et al* (1992) Development and phase I clinical testing of a conjugate vaccine against meningococcus A and C. *Vaccine* 10:691-698).

Meningococcus B remains a problem, however. This serotype currently is responsible for approximately 50% of total meningitis in the United States, Europe, and South America. The polysaccharide approach cannot be used because the menB capsular polysaccharide is a polymer of $\alpha(2-8)$ -linked *N*-acetyl neuraminic acid that is also present in mammalian tissue. This results in tolerance to the antigen; indeed, if an immune response were elicited, it would be anti-self, and therefore undesirable. In order to avoid induction of autoimmunity and to induce a protective immune response, the capsular polysaccharide has, for instance, been chemically modified substituting the *N*-acetyl groups with *N*-propionyl groups, leaving the specific antigenicity unaltered (Romero & Outschoorn (1994) Current status of Meningococcal group B vaccine candidates: capsular or non-capsular? *Clin Microbiol Rev* 7(4):559-575).

Alternative approaches to menB vaccines have used complex mixtures of outer membrane proteins (OMPs), containing either the OMPs alone, or OMPs enriched in porins, or deleted of the class 4

OMP's that are believed to induce antibodies that block bactericidal activity. This approach produces vaccines that are not well characterized. They are able to protect against the homologous strain, but are not effective at large where there are many antigenic variants of the outer membrane proteins. To overcome the antigenic variability, multivalent vaccines containing up to nine different porins have been constructed (eg. Poolman JT (1992) Development of a meningococcal vaccine. *Infect. Agents Dis.* 4:13-28). Additional proteins to be used in outer membrane vaccines have been the opa and ope proteins, but none of these approaches have been able to overcome the antigenic variability (eg. Ala'Aldeen & Borriello (1996) The meningococcal transferrin-binding proteins 1 and 2 are both surface exposed and generate bactericidal antibodies capable of killing homologous and heterologous strains. *Vaccine* 14(1):49-53).

A certain amount of sequence data is available for meningococcal and gonococcal genes and proteins (eg. EP-A-0467714, WO96/29412), but this is by no means complete. The provision of further sequences could provide an opportunity to identify secreted or surface-exposed proteins that are presumed targets for the immune system and which are not antigenically variable. For instance, some of the identified proteins could be components of efficacious vaccines against meningococcus B, some could be components of vaccines against all meningococcal serotypes, and others could be components of vaccines against all pathogenic *Neisseriae*.

THE INVENTION

The invention provides proteins comprising the Neisserial amino acid sequences disclosed in the examples. These sequences relate to *N.meningitidis* or *N.gonorrhoeae*.

It also provides proteins comprising sequences homologous (*ie.* having sequence identity) to the Neisserial amino acid sequences disclosed in the examples. Depending on the particular sequence, the degree of identity is preferably greater than 50% (eg. 65%, 80%, 90%, or more). These homologous proteins include mutants and allelic variants of the sequences disclosed in the examples. Typically, 50% identity or more between two proteins is considered to be an indication of functional equivalence. Identity between the proteins is preferably determined by the Smith-Waterman homology search algorithm as implemented in the MPSRCH program (Oxford Molecular), using an affine gap search with parameters *gap open penalty*=12 and *gap extension penalty*=1.

The invention further provides proteins comprising fragments of the Neisserial amino acid sequences disclosed in the examples. The fragments should comprise at least n consecutive amino acids from the sequences and, depending on the particular sequence, n is 7 or more (eg. 8, 10, 12, 14, 16, 18, 20 or more). Preferably the fragments comprise an epitope from the sequence.

- 5 The proteins of the invention can, of course, be prepared by various means (eg. recombinant expression, purification from cell culture, chemical synthesis *etc.*) and in various forms (eg. native, fusions *etc.*). They are preferably prepared in substantially pure or isolated form (*ie.* substantially free from other Neisserial or host cell proteins)

10 According to a further aspect, the invention provides antibodies which bind to these proteins. These may be polyclonal or monoclonal and may be produced by any suitable means.

According to a further aspect, the invention provides nucleic acid comprising the Neisserial nucleotide sequences disclosed in the examples. In addition, the invention provides nucleic acid comprising sequences homologous (*ie.* having sequence identity) to the Neisserial nucleotide sequences disclosed in the examples.

- 15 Furthermore, the invention provides nucleic acid which can hybridise to the Neisserial nucleic acid disclosed in the examples, preferably under "high stringency" conditions (eg. 65°C in a 0.1xSSC, 0.5% SDS solution).

20 Nucleic acid comprising fragments of these sequences are also provided. These should comprise at least n consecutive nucleotides from the Neisserial sequences and, depending on the particular sequence, n is 10 or more (eg 12, 14, 15, 18, 20, 25, 30, 35, 40 or more).

According to a further aspect, the invention provides nucleic acid encoding the proteins and protein fragments of the invention.

It should also be appreciated that the invention provides nucleic acid comprising sequences complementary to those described above (eg. for antisense or probing purposes).

Nucleic acid according to the invention can, of course, be prepared in many ways (*eg.* by chemical synthesis, from genomic or cDNA libraries, from the organism itself *etc.*) and can take various forms (*eg.* single stranded, double stranded, vectors, probes *etc.*).

In addition, the term "nucleic acid" includes DNA and RNA, and also their analogues, such as those containing modified backbones, and also peptide nucleic acids (PNA) *etc.*

According to a further aspect, the invention provides vectors comprising nucleotide sequences of the invention (*eg.* expression vectors) and host cells transformed with such vectors.

According to a further aspect, the invention provides compositions comprising protein, antibody, and/or nucleic acid according to the invention. These compositions may be suitable as vaccines, for instance, or as diagnostic reagents, or as immunogenic compositions.

The invention also provides nucleic acid, protein, or antibody according to the invention for use as medicaments (*eg.* as vaccines) or as diagnostic reagents. It also provides the use of nucleic acid, protein, or antibody according to the invention in the manufacture of: (i) a medicament for treating or preventing infection due to Neisserial bacteria; (ii) a diagnostic reagent for detecting the presence of Neisserial bacteria or of antibodies raised against Neisserial bacteria; and/or (iii) a reagent which can raise antibodies against Neisserial bacteria. Said Neisserial bacteria may be any species or strain (such as *N.gonorrhoeae*, or any strain of *N.meningitidis*, such as strain A, strain B or strain C).

The invention also provides a method of treating a patient, comprising administering to the patient a therapeutically effective amount of nucleic acid, protein, and/or antibody according to the invention.

According to further aspects, the invention provides various processes.

A process for producing proteins of the invention is provided, comprising the step of culturing a host cell according to the invention under conditions which induce protein expression.

A process for producing protein or nucleic acid of the invention is provided, wherein the the protein or nucleic acid is synthesised in part or in whole using chemical means.

A process for detecting polynucleotides of the invention is provided, comprising the steps of: (a) contacting a nucleic probe according to the invention with a biological sample under hybridizing conditions to form duplexes; and (b) detecting said duplexes.

A process for detecting proteins of the invention is provided, comprising the steps of: (a) contacting an antibody according to the invention with a biological sample under conditions suitable for the formation of an antibody-antigen complexes; and (b) detecting said complexes.

A summary of standard techniques and procedures which may be employed in order to perform the invention (eg. to utilise the disclosed sequences for vaccination or diagnostic purposes) follows. This summary is not a limitation on the invention but, rather, gives examples that may be used, but are not required.

General

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of molecular biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature eg. Sambrook *Molecular Cloning; A Laboratory Manual, Second Edition* (1989); *DNA Cloning, Volumes I and ii* (D.N. Glover ed. 1985); *Oligonucleotide Synthesis* (M.J. Gait ed, 1984); *Nucleic Acid Hybridization* (B.D. Hames & S.J. Higgins eds. 1984); *Transcription and Translation* (B.D. Hames & S.J. Higgins eds. 1984); *Animal Cell Culture* (R.I. Freshney ed. 1986); *Immobilized Cells and Enzymes* (IRL Press, 1986); B. Perbal, *A Practical Guide to Molecular Cloning* (1984); the *Methods in Enzymology* series (Academic Press, Inc.), especially volumes 154 & 155; *Gene Transfer Vectors for Mammalian Cells* (J.H. Miller and M.P. Calos eds. 1987, Cold Spring Harbor Laboratory); Mayer and Walker, eds. (1987), *Immunochemical Methods in Cell and Molecular Biology* (Academic Press, London); Scopes, (1987) *Protein Purification: Principles and Practice*, Second Edition (Springer-Verlag, N.Y.), and *Handbook of Experimental Immunology, Volumes I-IV* (D.M. Weir and C. C. Blackwell eds 1986).

Standard abbreviations for nucleotides and amino acids are used in this specification.

All publications, patents, and patent applications cited herein are incorporated in full by reference. In particular, the contents of UK patent applications 9723516.2, 9724190.5, 9724386.9, 9725158.1, 9726147.3, 9800759.4, and 9819016.8 are incorporated herein.

Definitions

A composition containing X is "substantially free of" Y when at least 85% by weight of the total X+Y in the composition is X. Preferably, X comprises at least about 90% by weight of the total of X+Y in the composition, more preferably at least about 95% or even 99% by weight.

- 5 The term "comprising" means "including" as well as "consisting" *eg.* a composition "comprising" X may consist exclusively of X or may include something additional to X, such as X+Y.

- A "conserved" *Neisseria* amino acid fragment or protein is one that is present in a particular Neisserial protein in at least x% of *Neisseria*. The value of x may be 50% or more, e.g., 66%, 75%, 80%, 90%, 95% or even 100% (i.e. the amino acid is found in the protein in question in all
10 *Neisseria*). In order to determine whether an amino acid is "conserved" in a particular Neisserial protein, it is necessary to compare that amino acid residue in the sequences of the protein in question from a plurality of different *Neisseria* (a reference population). The reference population may include a number of different *Neisseria* species or may include a single species. The reference
15 population may include a number of different serogroups of a particular species or a single serogroup. A preferred reference population consists of the 5 most common *Neisseria*. The term "heterologous" refers to two biological components that are not found together in nature. The components may be host cells, genes, or regulatory regions, such as promoters. Although the heterologous components are not found together in nature, they can function together, as when a promoter heterologous to a gene is operably linked to the gene. Another example is where a
20 Neisserial sequence is heterologous to a mouse host cell. A further examples would be two epitopes from the same or different proteins which have been assembled in a single protein in an arrangement not found in nature.

- An "origin of replication" is a polynucleotide sequence that initiates and regulates replication of polynucleotides, such as an expression vector. The origin of replication behaves as an autonomous
25 unit of polynucleotide replication within a cell, capable of replication under its own control. An origin of replication may be needed for a vector to replicate in a particular host cell. With certain origins of replication, an expression vector can be reproduced at a high copy number in the presence of the appropriate proteins within the cell. Examples of origins are the autonomously replicating sequences, which are effective in yeast; and the viral T-antigen, effective in COS-7
30 cells.

A "mutant" sequence is defined as DNA, RNA or amino acid sequence differing from but having sequence identity with the native or disclosed sequence. Depending on the particular sequence, the degree of sequence identity between the native or disclosed sequence and the mutant sequence is preferably greater than 50% (eg. 60%, 70%, 80%, 90%, 95%, 99% or more, calculated using the Smith-Waterman algorithm as described above). As used herein, an "allelic variant" of a nucleic acid molecule, or region, for which nucleic acid sequence is provided herein is a nucleic acid molecule, or region, that occurs essentially at the same locus in the genome of another or second isolate, and that, due to natural variation caused by, for example, mutation or recombination, has a similar but not identical nucleic acid sequence. A coding region allelic variant typically encodes a protein having similar activity to that of the protein encoded by the gene to which it is being compared. An allelic variant can also comprise an alteration in the 5' or 3' untranslated regions of the gene, such as in regulatory control regions (eg. see US patent 5,753,235).

Expression systems

The Neisserial nucleotide sequences can be expressed in a variety of different expression systems; for example those used with mammalian cells, baculoviruses, plants, bacteria, and yeast.

i. Mammalian Systems

Mammalian expression systems are known in the art. A mammalian promoter is any DNA sequence capable of binding mammalian RNA polymerase and initiating the downstream (3') transcription of a coding sequence (eg. structural gene) into mRNA. A promoter will have a transcription initiating region, which is usually placed proximal to the 5' end of the coding sequence, and a TATA box, usually located 25-30 base pairs (bp) upstream of the transcription initiation site. The TATA box is thought to direct RNA polymerase II to begin RNA synthesis at the correct site. A mammalian promoter will also contain an upstream promoter element, usually located within 100 to 200 bp upstream of the TATA box. An upstream promoter element determines the rate at which transcription is initiated and can act in either orientation [Sambrook et al. (1989) "Expression of Cloned Genes in Mammalian Cells." In *Molecular Cloning: A Laboratory Manual*, 2nd ed.].

Mammalian viral genes are often highly expressed and have a broad host range; therefore sequences encoding mammalian viral genes provide particularly useful promoter sequences. Examples include the SV40 early promoter, mouse mammary tumor virus LTR promoter, adenovirus major late

promoter (Ad MLP), and herpes simplex virus promoter. In addition, sequences derived from non-viral genes, such as the murine metallothionein gene, also provide useful promoter sequences. Expression may be either constitutive or regulated (inducible), depending on the promoter can be induced with glucocorticoid in hormone-responsive cells.

5 The presence of an enhancer element (enhancer), combined with the promoter elements described above, will usually increase expression levels. An enhancer is a regulatory DNA sequence that can stimulate transcription up to 1000-fold when linked to homologous or heterologous promoters, with synthesis beginning at the normal RNA start site. Enhancers are also active when they are placed upstream or downstream from the transcription initiation site, in either normal or flipped orientation, or at a distance of more than 1000 nucleotides from the promoter [Maniatis et al. (1987) *Science* 236:1237; Alberts et al. (1989) *Molecular Biology of the Cell*, 2nd ed.]. Enhancer elements derived from viruses may be particularly useful, because they usually have a broader host range. Examples include the SV40 early gene enhancer [Dijkema et al (1985) *EMBO J.* 4:761] and the enhancer/promoters derived from the long terminal repeat (LTR) of the Rous Sarcoma Virus [Gorman et al. (1982b) *Proc. Natl. Acad. Sci.* 79:6777] and from human cytomegalovirus [Boshart et al. (1985) *Cell* 41:521]. Additionally, some enhancers are regulatable and become active only in the presence of an inducer, such as a hormone or metal ion [Sassone-Corsi and Borelli (1986) *Trends Genet.* 2:215; Maniatis et al. (1987) *Science* 236:1237].

A DNA molecule may be expressed intracellularly in mammalian cells. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus of the recombinant protein will always be a methionine, which is encoded by the ATG start codon. If desired, the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide.

Alternatively, foreign proteins can also be secreted from the cell into the growth media by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provides for secretion of the foreign protein in mammalian cells. Preferably, there are processing sites encoded between the leader fragment and the foreign gene that can be cleaved either *in vivo* or *in vitro*. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell. The adenovirus tripartite leader is an example of a leader sequence that provides for secretion of a foreign protein in mammalian cells.

Usually, transcription termination and polyadenylation sequences recognized by mammalian cells are regulatory regions located 3' to the translation stop codon and thus, together with the promoter elements, flank the coding sequence. The 3' terminus of the mature mRNA is formed by site-specific post-transcriptional cleavage and polyadenylation [Birnstiel et al. (1985) *Cell* 41:349; Proudfoot and Whitelaw (1988) "Termination and 3' end processing of eukaryotic RNA. In *Transcription and splicing* (ed. B.D. Hames and D.M. Glover); Proudfoot (1989) *Trends Biochem. Sci.* 14:105]. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Examples of transcription terminator/polyadenylation signals include those derived from SV40 [Sambrook et al (1989) "Expression of cloned genes in cultured mammalian cells." In *Molecular Cloning: A Laboratory Manual*].

Usually, the above described components, comprising a promoter, polyadenylation signal, and transcription termination sequence are put together into expression constructs. Enhancers, introns with functional splice donor and acceptor sites, and leader sequences may also be included in an expression construct, if desired. Expression constructs are often maintained in a replicon, such as an extrachromosomal element (eg. plasmids) capable of stable maintenance in a host, such as mammalian cells or bacteria. Mammalian replication systems include those derived from animal viruses, which require trans-acting factors to replicate. For example, plasmids containing the replication systems of papovaviruses, such as SV40 [Gluzman (1981) *Cell* 23:175] or polyomavirus, replicate to extremely high copy number in the presence of the appropriate viral T antigen. Additional examples of mammalian replicons include those derived from bovine papillomavirus and Epstein-Barr virus. Additionally, the replicon may have two replicon systems, thus allowing it to be maintained, for example, in mammalian cells for expression and in a prokaryotic host for cloning and amplification. Examples of such mammalian-bacteria shuttle vectors include pMT2 [Kaufman et al. (1989) *Mol. Cell. Biol.* 9:946] and pHEBO [Shimizu et al. (1986) *Mol. Cell. Biol.* 6:1074].

The transformation procedure used depends upon the host to be transformed. Methods for introduction of heterologous polynucleotides into mammalian cells are known in the art and include dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the DNA into nuclei.

Mammalian cell lines available as hosts for expression are known in the art and include many immortalized cell lines available from the American Type Culture Collection (ATCC), including but not limited to, Chinese hamster ovary (CHO) cells, HeLa cells, baby hamster kidney (BHK) cells, monkey kidney cells (COS), human hepatocellular carcinoma cells (eg. Hep G2), and a number of other cell lines.

ii. Baculovirus Systems

The polynucleotide encoding the protein can also be inserted into a suitable insect expression vector, and is operably linked to the control elements within that vector. Vector construction employs techniques which are known in the art. Generally, the components of the expression system include a transfer vector, usually a bacterial plasmid, which contains both a fragment of the baculovirus genome, and a convenient restriction site for insertion of the heterologous gene or genes to be expressed; a wild type baculovirus with a sequence homologous to the baculovirus-specific fragment in the transfer vector (this allows for the homologous recombination of the heterologous gene in to the baculovirus genome); and appropriate insect host cells and growth media.

After inserting the DNA sequence encoding the protein into the transfer vector, the vector and the wild type viral genome are transfected into an insect host cell where the vector and viral genome are allowed to recombine. The packaged recombinant virus is expressed and recombinant plaques are identified and purified. Materials and methods for baculovirus/insect cell expression systems are commercially available in kit form from, *inter alia*, Invitrogen, San Diego CA ("MaxBac" kit). These techniques are generally known to those skilled in the art and fully described in Summers and Smith, *Texas Agricultural Experiment Station Bulletin No. 1555* (1987) (hereinafter "Summers and Smith").

Prior to inserting the DNA sequence encoding the protein into the baculovirus genome, the above described components, comprising a promoter, leader (if desired), coding sequence of interest, and transcription termination sequence, are usually assembled into an intermediate transplacement construct (transfer vector). This construct may contain a single gene and operably linked regulatory elements; multiple genes, each with its own set of operably linked regulatory elements; or multiple genes, regulated by the same set of regulatory elements. Intermediate transplacement constructs are often maintained in a replicon, such as an extrachromosomal element (eg. plasmids)

capable of stable maintenance in a host, such as a bacterium. The replicon will have a replication system, thus allowing it to be maintained in a suitable host for cloning and amplification.

Currently, the most commonly used transfer vector for introducing foreign genes into AcNPV is pAc373. Many other vectors, known to those of skill in the art, have also been designed. These include, for example, pVL985 (which alters the polyhedrin start codon from ATG to ATT, and which introduces a BamHI cloning site 32 basepairs downstream from the ATT; see Luckow and Summers, *Virology* (1989) 17:31.

The plasmid usually also contains the polyhedrin polyadenylation signal (Miller et al. (1988) *Ann. Rev. Microbiol.*, 42:177) and a prokaryotic ampicillin-resistance (*amp*) gene and origin of replication for selection and propagation in *E. coli*.

Baculovirus transfer vectors usually contain a baculovirus promoter. A baculovirus promoter is any DNA sequence capable of binding a baculovirus RNA polymerase and initiating the downstream (5' to 3') transcription of a coding sequence (eg. structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site and a transcription initiation site. A baculovirus transfer vector may also have a second domain called an enhancer, which, if present, is usually distal to the structural gene. Expression may be either regulated or constitutive.

Structural genes, abundantly transcribed at late times in a viral infection cycle, provide particularly useful promoter sequences. Examples include sequences derived from the gene encoding the viral polyhedrin protein, Friesen et al., (1986) "The Regulation of Baculovirus Gene Expression," in: *The Molecular Biology of Baculoviruses* (ed. Walter Doerfler); EPO Publ. Nos. 127 839 and 155 476; and the gene encoding the p10 protein, Vlsek et al., (1988), *J. Gen. Virol.* 69:765.

DNA encoding suitable signal sequences can be derived from genes for secreted insect or baculovirus proteins, such as the baculovirus polyhedrin gene (Carbonell et al. (1988) *Gene*, 73:409). Alternatively, since the signals for mammalian cell posttranslational modifications (such as signal peptide cleavage, proteolytic cleavage, and phosphorylation) appear to be recognized by insect cells, and the signals required for secretion and nuclear accumulation also appear to be conserved between the invertebrate cells and vertebrate cells, leaders of non-insect origin, such as

those derived from genes encoding human α -interferon, Maeda et al., (1985), *Nature* 315:592; human gastrin-releasing peptide, Lebacqz-Verheyden et al., (1988), *Molec. Cell. Biol.* 8:3129; human IL-2, Smith et al., (1985) *Proc. Nat'l Acad. Sci. USA*, 82:8404; mouse IL-3, (Miyajima et al., (1987) *Gene* 58:273; and human glucocerebrosidase, Martin et al. (1988) *DNA*, 7:99, can also
5 be used to provide for secretion in insects.

A recombinant polypeptide or polyprotein may be expressed intracellularly or, if it is expressed with the proper regulatory sequences, it can be secreted. Good intracellular expression of nonfused foreign proteins usually requires heterologous genes that ideally have a short leader sequence containing suitable translation initiation signals preceding an ATG start signal. If desired,
10 methionine at the N-terminus may be cleaved from the mature protein by *in vitro* incubation with cyanogen bromide.

Alternatively, recombinant polyproteins or proteins which are not naturally secreted can be secreted from the insect cell by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provides for secretion of the foreign protein in insects. The
15 leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the translocation of the protein into the endoplasmic reticulum.

After insertion of the DNA sequence and/or the gene encoding the expression product precursor of the protein, an insect cell host is co-transformed with the heterologous DNA of the transfer vector and the genomic DNA of wild type baculovirus -- usually by co-transfection. The promoter
20 and transcription termination sequence of the construct will usually comprise a 2-5kb section of the baculovirus genome. Methods for introducing heterologous DNA into the desired site in the baculovirus virus are known in the art. (See Summers and Smith *supra*; Ju et al. (1987); Smith et al., *Mol. Cell. Biol.* (1983) 3:2156; and Luckow and Summers (1989)). For example, the insertion can be into a gene such as the polyhedrin gene, by homologous double crossover recombination;
25 insertion can also be into a restriction enzyme site engineered into the desired baculovirus gene. Miller et al., (1989), *Bioessays* 4:91. The DNA sequence, when cloned in place of the polyhedrin gene in the expression vector, is flanked both 5' and 3' by polyhedrin-specific sequences and is positioned downstream of the polyhedrin promoter.

The newly formed baculovirus expression vector is subsequently packaged into an infectious recombinant baculovirus. Homologous recombination occurs at low frequency (between about 1% and about 5%); thus, the majority of the virus produced after cotransfection is still wild-type virus. Therefore, a method is necessary to identify recombinant viruses. An advantage of the expression system is a visual screen allowing recombinant viruses to be distinguished. The polyhedrin protein, which is produced by the native virus, is produced at very high levels in the nuclei of infected cells at late times after viral infection. Accumulated polyhedrin protein forms occlusion bodies that also contain embedded particles. These occlusion bodies, up to 15 μm in size, are highly refractile, giving them a bright shiny appearance that is readily visualized under the light microscope. Cells infected with recombinant viruses lack occlusion bodies. To distinguish recombinant virus from wild-type virus, the transfection supernatant is plaqued onto a monolayer of insect cells by techniques known to those skilled in the art. Namely, the plaques are screened under the light microscope for the presence (indicative of wild-type virus) or absence (indicative of recombinant virus) of occlusion bodies. "Current Protocols in Microbiology" Vol. 2 (Ausubel et al. eds) at 16.8 (Supp. 10, 1990); Summers and Smith, *supra*; Miller et al. (1989).

Recombinant baculovirus expression vectors have been developed for infection into several insect cells. For example, recombinant baculoviruses have been developed for, *inter alia*: *Aedes aegypti*, *Autographa californica*, *Bombyx mori*, *Drosophila melanogaster*, *Spodoptera frugiperda*, and *Trichoplusia ni* (WO 89/046699; Carbonell et al., (1985) *J. Virol.* 56:153; Wright (1986) *Nature* 321:718; Smith et al., (1983) *Mol. Cell. Biol.* 3:2156; and see generally, Fraser, et al. (1989) *In Vitro Cell. Dev. Biol.* 25:225).

Cells and cell culture media are commercially available for both direct and fusion expression of heterologous polypeptides in a baculovirus/expression system; cell culture technology is generally known to those skilled in the art. See, eg. Summers and Smith *supra*.

The modified insect cells may then be grown in an appropriate nutrient medium, which allows for stable maintenance of the plasmid(s) present in the modified insect host. Where the expression product gene is under inducible control, the host may be grown to high density, and expression induced. Alternatively, where expression is constitutive, the product will be continuously expressed into the medium and the nutrient medium must be continuously circulated, while removing the product of interest and augmenting depleted nutrients. The product may be purified by such

techniques as chromatography, eg. HPLC, affinity chromatography, ion exchange chromatography, etc.; electrophoresis; density gradient centrifugation; solvent extraction, or the like. As appropriate, the product may be further purified, as required, so as to remove substantially any insect proteins which are also secreted in the medium or result from lysis of insect cells, so as to provide a product which is at least substantially free of host debris, eg. proteins, lipids and polysaccharides.

In order to obtain protein expression, recombinant host cells derived from the transformants are incubated under conditions which allow expression of the recombinant protein encoding sequence. These conditions will vary, dependent upon the host cell selected. However, the conditions are readily ascertainable to those of ordinary skill in the art, based upon what is known in the art.

iii. Plant Systems

There are many plant cell culture and whole plant genetic expression systems known in the art. Exemplary plant cellular genetic expression systems include those described in patents, such as: US 5,693,506; US 5,659,122; and US 5,608,143. Additional examples of genetic expression in plant cell culture has been described by Zenk, *Phytochemistry* 30:3861-3863 (1991). Descriptions of plant protein signal peptides may be found in addition to the references described above in Vaulcombe et al., *Mol. Gen. Genet.* 209:33-40 (1987); Chandler et al., *Plant Molecular Biology* 3:407-418 (1984); Rogers, *J. Biol. Chem.* 260:3731-3738 (1985); Rothstein et al., *Gene* 55:353-356 (1987); Whittier et al., *Nucleic Acids Research* 15:2515-2535 (1987); Wirsal et al., *Molecular Microbiology* 3:3-14 (1989); Yu et al., *Gene* 122:247-253 (1992). A description of the regulation of plant gene expression by the phytohormone, gibberellic acid and secreted enzymes induced by gibberellic acid can be found in R.L. Jones and J. MacMillin, *Gibberellins: in: Advanced Plant Physiology*, Malcolm B. Wilkins, ed., 1984 Pitman Publishing Limited, London, pp. 21-52. References that describe other metabolically-regulated genes: Sheen, *Plant Cell*, 2:1027-1038(1990); Maas et al., *EMBO J.* 9:3447-3452 (1990); Benkel and Hickey, *Proc. Natl. Acad. Sci.* 84:1337-1339 (1987)

Typically, using techniques known in the art, a desired polynucleotide sequence is inserted into an expression cassette comprising genetic regulatory elements designed for operation in plants. The expression cassette is inserted into a desired expression vector with companion sequences upstream and downstream from the expression cassette suitable for expression in a plant host. The companion sequences will be of plasmid or viral origin and provide necessary characteristics to the

vector to permit the vectors to move DNA from an original cloning host, such as bacteria, to the desired plant host. The basic bacterial/plant vector construct will preferably provide a broad host range prokaryote replication origin; a prokaryote selectable marker; and, for *Agrobacterium* transformations, T DNA sequences for *Agrobacterium*-mediated transfer to plant chromosomes.

- 5 Where the heterologous gene is not readily amenable to detection, the construct will preferably also have a selectable marker gene suitable for determining if a plant cell has been transformed. A general review of suitable markers, for example for the members of the grass family, is found in Wilmink and Dons, 1993, *Plant Mol. Biol. Repr.*, 11(2):165-185.

- 10 Sequences suitable for permitting integration of the heterologous sequence into the plant genome are also recommended. These might include transposon sequences and the like for homologous recombination as well as Ti sequences which permit random insertion of a heterologous expression cassette into a plant genome. Suitable prokaryote selectable markers include resistance toward antibiotics such as ampicillin or tetracycline. Other DNA sequences encoding additional functions may also be present in the vector, as is known in the art.

- 15 The nucleic acid molecules of the subject invention may be included into an expression cassette for expression of the protein(s) of interest. Usually, there will be only one expression cassette, although two or more are feasible. The recombinant expression cassette will contain in addition to the heterologous protein encoding sequence the following elements, a promoter region, plant 5' untranslated sequences, initiation codon depending upon whether or not the structural gene comes
20 equipped with one, and a transcription and translation termination sequence. Unique restriction enzyme sites at the 5' and 3' ends of the cassette allow for easy insertion into a pre-existing vector.

- A heterologous coding sequence may be for any protein relating to the present invention. The sequence encoding the protein of interest will encode a signal peptide which allows processing and translocation of the protein, as appropriate, and will usually lack any sequence which might result
25 in the binding of the desired protein of the invention to a membrane. Since, for the most part, the transcriptional initiation region will be for a gene which is expressed and translocated during germination, by employing the signal peptide which provides for translocation, one may also provide for translocation of the protein of interest. In this way, the protein(s) of interest will be translocated from the cells in which they are expressed and may be efficiently harvested. Typically
30 secretion in seeds are across the aleurone or scutellar epithelium layer into the endosperm of the

seed. While it is not required that the protein be secreted from the cells in which the protein is produced, this facilitates the isolation and purification of the recombinant protein.

Since the ultimate expression of the desired gene product will be in a eucaryotic cell it is desirable to determine whether any portion of the cloned gene contains sequences which will be processed out as introns by the host's spliceosome machinery. If so, site-directed mutagenesis of the "intron" region may be conducted to prevent losing a portion of the genetic message as a false intron code, Reed and Maniatis, *Cell* 41:95-105, 1985.

The vector can be microinjected directly into plant cells by use of micropipettes to mechanically transfer the recombinant DNA. Crossway, *Mol. Gen. Genet.*, 202:179-185, 1985. The genetic material may also be transferred into the plant cell by using polyethylene glycol, Krens, et al., *Nature*, 296, 72-74, 1982. Another method of introduction of nucleic acid segments is high velocity ballistic penetration by small particles with the nucleic acid either within the matrix of small beads or particles, or on the surface, Klein, et al., *Nature*, 327, 70-73, 1987 and Knudsen and Muller, 1991, *Planta*, 185:330-336 teaching particle bombardment of barley endosperm to create transgenic barley. Yet another method of introduction would be fusion of protoplasts with other entities, either minicells, cells, lysosomes or other fusible lipid-surfaced bodies, Fraley, et al., *Proc. Natl. Acad. Sci. USA*, 79, 1859-1863, 1982.

The vector may also be introduced into the plant cells by electroporation. (Fromm et al., *Proc. Natl. Acad. Sci. USA* 82:5824, 1985). In this technique, plant protoplasts are electroporated in the presence of plasmids containing the gene construct. Electrical impulses of high field strength reversibly permeabilize biomembranes allowing the introduction of the plasmids. Electroporated plant protoplasts reform the cell wall, divide, and form plant callus.

All plants from which protoplasts can be isolated and cultured to give whole regenerated plants can be transformed by the present invention so that whole plants are recovered which contain the transferred gene. It is known that practically all plants can be regenerated from cultured cells or tissues, including but not limited to all major species of sugarcane, sugar beet, cotton, fruit and other trees, legumes and vegetables. Some suitable plants include, for example, species from the genera *Fragaria*, *Lotus*, *Medicago*, *Onobrychis*, *Trifolium*, *Trigonella*, *Vigna*, *Citrus*, *Linum*, *Geranium*, *Manihot*, *Daucus*, *Arabidopsis*, *Brassica*, *Raphanus*, *Sinapis*, *Atropa*, *Capsicum*, *Datura*, *Hyoscyamus*, *Lycopersion*, *Nicotiana*, *Solanum*, *Petunia*, *Digitalis*, *Majorana*, *Cichorium*,

Helianthus, Lactuca, Bromus, Asparagus, Antirrhinum, Hererocallis, Nemesia, Pelargonium, Panicum, Pennisetum, Ranunculus, Senecio, Salpiglossis, Cucumis, Browaalia, Glycine, Lolium, Zea, Triticum, Sorghum, and Datura.

Means for regeneration vary from species to species of plants, but generally a suspension of transformed protoplasts containing copies of the heterologous gene is first provided. Callus tissue is formed and shoots may be induced from callus and subsequently rooted. Alternatively, embryo formation can be induced from the protoplast suspension. These embryos germinate as natural embryos to form plants. The culture media will generally contain various amino acids and hormones, such as auxin and cytokinins. It is also advantageous to add glutamic acid and proline to the medium, especially for such species as corn and alfalfa. Shoots and roots normally develop simultaneously. Efficient regeneration will depend on the medium, on the genotype, and on the history of the culture. If these three variables are controlled, then regeneration is fully reproducible and repeatable.

In some plant cell culture systems, the desired protein of the invention may be excreted or alternatively, the protein may be extracted from the whole plant. Where the desired protein of the invention is secreted into the medium, it may be collected. Alternatively, the embryos and embryoless-half seeds or other plant tissue may be mechanically disrupted to release any secreted protein between cells and tissues. The mixture may be suspended in a buffer solution to retrieve soluble proteins. Conventional protein isolation and purification methods will be then used to purify the recombinant protein. Parameters of time, temperature pH, oxygen, and volumes will be adjusted through routine methods to optimize expression and recovery of heterologous protein.

iv. Bacterial Systems

Bacterial expression techniques are known in the art. A bacterial promoter is any DNA sequence capable of binding bacterial RNA polymerase and initiating the downstream (3') transcription of a coding sequence (eg. structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site and a transcription initiation site. A bacterial promoter may also have a second domain called an operator, that may overlap an adjacent RNA polymerase binding site at which RNA synthesis begins. The operator permits negative regulated (inducible) transcription, as a gene repressor protein may bind the operator and

thereby inhibit transcription of a specific gene. Constitutive expression may occur in the absence of negative regulatory elements, such as the operator. In addition, positive regulation may be achieved by a gene activator protein binding sequence, which, if present is usually proximal (5') to the RNA polymerase binding sequence. An example of a gene activator protein is the catabolite activator protein (CAP), which helps initiate transcription of the lac operon in *Escherichia coli* (*E. coli*) [Raibaud *et al.* (1984) *Annu. Rev. Genet.* 18:173]. Regulated expression may therefore be either positive or negative, thereby either enhancing or reducing transcription.

Sequences encoding metabolic pathway enzymes provide particularly useful promoter sequences. Examples include promoter sequences derived from sugar metabolizing enzymes, such as galactose, lactose (*lac*) [Chang *et al.* (1977) *Nature* 198:1056], and maltose. Additional examples include promoter sequences derived from biosynthetic enzymes such as tryptophan (*trp*) [Goeddel *et al.* (1980) *Nuc. Acids Res.* 8:4057; Yelverton *et al.* (1981) *Nucl. Acids Res.* 9:731; US patent 4,738,921; EP-A-0036776 and EP-A-0121775]. The *g*-laotamase (*bla*) promoter system [Weissmann (1981) "The cloning of interferon and other mistakes." In *Interferon 3* (ed. I. Gresser)], bacteriophage lambda PL [Shimatake *et al.* (1981) *Nature* 292:128] and T5 [US patent 4,689,406] promoter systems also provide useful promoter sequences.

In addition, synthetic promoters which do not occur in nature also function as bacterial promoters. For example, transcription activation sequences of one bacterial or bacteriophage promoter may be joined with the operon sequences of another bacterial or bacteriophage promoter, creating a synthetic hybrid promoter [US patent 4,551,433]. For example, the *tac* promoter is a hybrid *trp-lac* promoter comprised of both *trp* promoter and *lac* operon sequences that is regulated by the *lac* repressor [Amann *et al.* (1983) *Gene* 25:167; de Boer *et al.* (1983) *Proc. Natl. Acad. Sci.* 80:21]. Furthermore, a bacterial promoter can include naturally occurring promoters of non-bacterial origin that have the ability to bind bacterial RNA polymerase and initiate transcription. A naturally occurring promoter of non-bacterial origin can also be coupled with a compatible RNA polymerase to produce high levels of expression of some genes in prokaryotes. The bacteriophage T7 RNA polymerase/promoter system is an example of a coupled promoter system [Studier *et al.* (1986) *J. Mol. Biol.* 189:113; Tabor *et al.* (1985) *Proc Natl. Acad. Sci.* 82:1074]. In addition, a hybrid promoter can also be comprised of a bacteriophage promoter and an *E. coli* operator region (EPO-A-0 267 851).

In addition to a functioning promoter sequence, an efficient ribosome binding site is also useful for the expression of foreign genes in prokaryotes. In *E. coli*, the ribosome binding site is called the Shine-Dalgarno (SD) sequence and includes an initiation codon (ATG) and a sequence 3-9 nucleotides in length located 3-11 nucleotides upstream of the initiation codon [Shine *et al.* (1975) *Nature* 254:34]. The SD sequence is thought to promote binding of mRNA to the ribosome by the pairing of bases between the SD sequence and the 3' end of *E. coli* 16S rRNA [Steitz *et al.* (1979) "Genetic signals and nucleotide sequences in messenger RNA." In *Biological Regulation and Development: Gene Expression* (ed. R.F. Goldberg)]. To express eukaryotic genes and prokaryotic genes with weak ribosome-binding site [Sambrook *et al.* (1989) "Expression of cloned genes in *Escherichia coli*." In *Molecular Cloning: A Laboratory Manual*].

A DNA molecule may be expressed intracellularly. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus will always be a methionine, which is encoded by the ATG start codon. If desired, methionine at the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide or by either *in vivo* on *in vitro* incubation with a bacterial methionine N-terminal peptidase (EPO-A-0 219 237).

Fusion proteins provide an alternative to direct expression. Usually, a DNA sequence encoding the N-terminal portion of an endogenous bacterial protein, or other stable protein, is fused to the 5' end of heterologous coding sequences. Upon expression, this construct will provide a fusion of the two amino acid sequences. For example, the bacteriophage lambda cell gene can be linked at the 5' terminus of a foreign gene and expressed in bacteria. The resulting fusion protein preferably retains a site for a processing enzyme (factor Xa) to cleave the bacteriophage protein from the foreign gene [Nagai *et al.* (1984) *Nature* 309:810]. Fusion proteins can also be made with sequences from the *lacZ* [Jia *et al.* (1987) *Gene* 60:197], *trpE* [Allen *et al.* (1987) *J. Biotechnol.* 5:93; Makoff *et al.* (1989) *J. Gen. Microbiol.* 135:11], and *Chey* [EP-A-0 324 647] genes. The DNA sequence at the junction of the two amino acid sequences may or may not encode a cleavable site. Another example is a ubiquitin fusion protein. Such a fusion protein is made with the ubiquitin region that preferably retains a site for a processing enzyme (*eg.* ubiquitin specific processing-protease) to cleave the ubiquitin from the foreign protein. Through this method, native foreign protein can be isolated [Miller *et al.* (1989) *Bio/Technology* 7:698].

Alternatively, foreign proteins can also be secreted from the cell by creating chimeric DNA molecules that encode a fusion protein comprised of a signal peptide sequence fragment that provides for secretion of the foreign protein in bacteria [US patent 4,336,336]. The signal sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell. The protein is either secreted into the growth media (gram-positive bacteria) or into the periplasmic space, located between the inner and outer membrane of the cell (gram-negative bacteria). Preferably there are processing sites, which can be cleaved either *in vivo* or *in vitro* encoded between the signal peptide fragment and the foreign gene.

DNA encoding suitable signal sequences can be derived from genes for secreted bacterial proteins, such as the *E. coli* outer membrane protein gene (*ompA*) [Masui *et al.* (1983), in: *Experimental Manipulation of Gene Expression*; Ghrayeb *et al.* (1984) *EMBO J.* 3:2437] and the *E. coli* alkaline phosphatase signal sequence (*phoA*) [Oka *et al.* (1985) *Proc. Natl. Acad. Sci.* 82:7212]. As an additional example, the signal sequence of the alpha-amylase gene from various *Bacillus* strains can be used to secrete heterologous proteins from *B. subtilis* [Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA* 79:5582; EP-A-0 244 042].

Usually, transcription termination sequences recognized by bacteria are regulatory regions located 3' to the translation stop codon, and thus together with the promoter flank the coding sequence. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Transcription termination sequences frequently include DNA sequences of about 50 nucleotides capable of forming stem loop structures that aid in terminating transcription. Examples include transcription termination sequences derived from genes with strong promoters, such as the *trp* gene in *E. coli* as well as other biosynthetic genes.

Usually, the above described components, comprising a promoter, signal sequence (if desired), coding sequence of interest, and transcription termination sequence, are put together into expression constructs. Expression constructs are often maintained in a replicon, such as an extrachromosomal element (eg. plasmids) capable of stable maintenance in a host, such as bacteria. The replicon will have a replication system, thus allowing it to be maintained in a prokaryotic host either for expression or for cloning and amplification. In addition, a replicon may be either a high or low copy number plasmid. A high copy number plasmid will generally have a copy number ranging from about 5 to about 200, and usually about 10 to about 150. A host containing a high copy

number plasmid will preferably contain at least about 10, and more preferably at least about 20 plasmids. Either a high or low copy number vector may be selected, depending upon the effect of the vector and the foreign protein on the host.

Alternatively, the expression constructs can be integrated into the bacterial genome with an integrating vector. Integrating vectors usually contain at least one sequence homologous to the bacterial chromosome that allows the vector to integrate. Integrations appear to result from recombinations between homologous DNA in the vector and the bacterial chromosome. For example, integrating vectors constructed with DNA from various *Bacillus* strains integrate into the *Bacillus* chromosome (EP-A- 0 127 328). Integrating vectors may also be comprised of bacteriophage or transposon sequences.

Usually, extrachromosomal and integrating expression constructs may contain selectable markers to allow for the selection of bacterial strains that have been transformed. Selectable markers can be expressed in the bacterial host and may include genes which render bacteria resistant to drugs such as ampicillin, chloramphenicol, erythromycin, kanamycin (neomycin), and tetracycline [Davies *et al.* (1978) *Annu. Rev. Microbiol.* 32:469]. Selectable markers may also include biosynthetic genes, such as those in the histidine, tryptophan, and leucine biosynthetic pathways.

Alternatively, some of the above described components can be put together in transformation vectors. Transformation vectors are usually comprised of a selectable market that is either maintained in a replicon or developed into an integrating vector, as described above.

Expression and transformation vectors, either extra-chromosomal replicons or integrating vectors, have been developed for transformation into many bacteria. For example, expression vectors have been developed for, *inter alia*, the following bacteria: *Bacillus subtilis* [Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA* 79:5582; EP-A-0 036 259 and EP-A-0 063 953; WO 84/04541], *Escherichia coli* [Shimatake *et al.* (1981) *Nature* 292:128; Amann *et al.* (1985) *Gene* 40:183; Studier *et al.* (1986) *J. Mol. Biol.* 189:113; EP-A-0 036 776, EP-A-0 136 829 and EP-A-0 136 907], *Streptococcus cremoris* [Powell *et al.* (1988) *Appl. Environ. Microbiol.* 54:655]; *Streptococcus lividans* [Powell *et al.* (1988) *Appl. Environ. Microbiol.* 54:655], *Streptomyces lividans* [US patent 4,745,056].

Methods of introducing exogenous DNA into bacterial hosts are well-known in the art, and usually include either the transformation of bacteria treated with CaCl₂ or other agents, such as divalent cations and DMSO. DNA can also be introduced into bacterial cells by electroporation. Transformation procedures usually vary with the bacterial species to be transformed. See eg. [Masson *et al.* (1989) *FEMS Microbiol. Lett.* 60:273; Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA* 79:5582; EP-A-0 036 259 and EP-A-0 063 953; WO 84/04541, *Bacillus*], [Miller *et al.* (1988) *Proc. Natl. Acad. Sci.* 85:856; Wang *et al.* (1990) *J. Bacteriol.* 172:949, *Campylobacter*], [Cohen *et al.* (1973) *Proc. Natl. Acad. Sci.* 69:2110; Dower *et al.* (1988) *Nucleic Acids Res.* 16:6127; Kushner (1978) "An improved method for transformation of *Escherichia coli* with ColE1-derived plasmids. In *Genetic Engineering: Proceedings of the International Symposium on Genetic Engineering* (eds. H.W. Boyer and S. Nicosia); Mandel *et al.* (1970) *J. Mol. Biol.* 53:159; Taketo (1988) *Biochim. Biophys. Acta* 949:318; *Escherichia*], [Chassy *et al.* (1987) *FEMS Microbiol. Lett.* 44:173 *Lactobacillus*]; [Fiedler *et al.* (1988) *Anal. Biochem* 170:38, *Pseudomonas*]; [Augustin *et al.* (1990) *FEMS Microbiol. Lett.* 66:203, *Staphylococcus*], [Barany *et al.* (1980) *J. Bacteriol.* 144:698; Harlander (1987) "Transformation of *Streptococcus lactis* by electroporation, in: *Streptococcal Genetics* (ed. J. Ferretti and R. Curtiss III); Perry *et al.* (1981) *Infect. Immun.* 32:1295; Powell *et al.* (1988) *Appl. Environ. Microbiol.* 54:655; Somkuti *et al.* (1987) *Proc. 4th Evr. Cong. Biotechnology* 1:412, *Streptococcus*].

v. Yeast Expression

Yeast expression systems are also known to one of ordinary skill in the art. A yeast promoter is any DNA sequence capable of binding yeast RNA polymerase and initiating the downstream (3') transcription of a coding sequence (eg. structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site (the "TATA Box") and a transcription initiation site. A yeast promoter may also have a second domain called an upstream activator sequence (UAS), which, if present, is usually distal to the structural gene. The UAS permits regulated (inducible) expression. Constitutive expression occurs in the absence of a UAS. Regulated expression may be either positive or negative, thereby either enhancing or reducing transcription.

Yeast is a fermenting organism with an active metabolic pathway, therefore sequences encoding enzymes in the metabolic pathway provide particularly useful promoter sequences. Examples

include alcohol dehydrogenase (ADH) (EP-A-0 284 044), enolase, glucokinase, glucose-6-phosphate isomerase, glyceraldehyde-3-phosphate-dehydrogenase (GAP or GAPDH), hexokinase, phosphofructokinase, 3-phosphoglycerate mutase, and pyruvate kinase (PyK) (EPO-A-0 329 203). The yeast *PHO5* gene, encoding acid phosphatase, also provides useful promoter sequences [Myanohara *et al.* (1983) *Proc. Natl. Acad. Sci. USA* 80:1].

In addition, synthetic promoters which do not occur in nature also function as yeast promoters. For example, UAS sequences of one yeast promoter may be joined with the transcription activation region of another yeast promoter, creating a synthetic hybrid promoter. Examples of such hybrid promoters include the ADH regulatory sequence linked to the GAP transcription activation region (US Patent Nos. 4,876,197 and 4,880,734). Other examples of hybrid promoters include promoters which consist of the regulatory sequences of either the *ADH2*, *GAL4*, *GAL10*, OR *PHO5* genes, combined with the transcriptional activation region of a glycolytic enzyme gene such as GAP or PyK (EP-A-0 164 556). Furthermore, a yeast promoter can include naturally occurring promoters of non-yeast origin that have the ability to bind yeast RNA polymerase and initiate transcription. Examples of such promoters include, *inter alia*, [Cohen *et al.* (1980) *Proc. Natl. Acad. Sci. USA* 77:1078; Henikoff *et al.* (1981) *Nature* 283:835; Hollenberg *et al.* (1981) *Curr. Topics Microbiol. Immunol.* 96:119; Hollenberg *et al.* (1979) "The Expression of Bacterial Antibiotic Resistance Genes in the Yeast *Saccharomyces cerevisiae*," in: *Plasmids of Medical, Environmental and Commercial Importance* (eds. K.N. Timmis and A. Puhler); Mercerau-Puigalon *et al.* (1980) *Gene* 11:163; Panthier *et al.* (1980) *Curr. Genet.* 2:109;].

A DNA molecule may be expressed intracellularly in yeast. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus of the recombinant protein will always be a methionine, which is encoded by the ATG start codon. If desired, methionine at the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide.

Fusion proteins provide an alternative for yeast expression systems, as well as in mammalian, baculovirus, and bacterial expression systems. Usually, a DNA sequence encoding the N-terminal portion of an endogenous yeast protein, or other stable protein, is fused to the 5' end of heterologous coding sequences. Upon expression, this construct will provide a fusion of the two amino acid sequences. For example, the yeast or human superoxide dismutase (SOD) gene, can be

linked at the 5' terminus of a foreign gene and expressed in yeast. The DNA sequence at the junction of the two amino acid sequences may or may not encode a cleavable site. See *eg.* EP-A-0 196 056. Another example is a ubiquitin fusion protein. Such a fusion protein is made with the ubiquitin region that preferably retains a site for a processing enzyme (*eg.* ubiquitin-specific processing protease) to cleave the ubiquitin from the foreign protein. Through this method, therefore, native foreign protein can be isolated (*eg.* WO88/024066).

Alternatively, foreign proteins can also be secreted from the cell into the growth media by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provide for secretion in yeast of the foreign protein. Preferably, there are processing sites encoded between the leader fragment and the foreign gene that can be cleaved either *in vivo* or *in vitro*. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell.

DNA encoding suitable signal sequences can be derived from genes for secreted yeast proteins, such as the yeast invertase gene (EP-A-0 012 873; JPO. 62,096,086) and the A-factor gene (US patent 4,588,684). Alternatively, leaders of non-yeast origin, such as an interferon leader, exist that also provide for secretion in yeast (EP-A-0 060 057).

A preferred class of secretion leaders are those that employ a fragment of the yeast alpha-factor gene, which contains both a "pre" signal sequence, and a "pro" region. The types of alpha-factor fragments that can be employed include the full-length pre-pro alpha factor leader (about 83 amino acid residues) as well as truncated alpha-factor leaders (usually about 25 to about 50 amino acid residues) (US Patents 4,546,083 and 4,870,008; EP-A-0 324 274). Additional leaders employing an alpha-factor leader fragment that provides for secretion include hybrid alpha-factor leaders made with a presequence of a first yeast, but a pro-region from a second yeast alphafactor. (*eg.* see WO 89/02463.)

Usually, transcription termination sequences recognized by yeast are regulatory regions located 3' to the translation stop codon, and thus together with the promoter flank the coding sequence. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Examples of transcription terminator sequence and other yeast-recognized termination sequences, such as those coding for glycolytic enzymes.

Usually, the above described components, comprising a promoter, leader (if desired), coding sequence of interest, and transcription termination sequence, are put together into expression constructs. Expression constructs are often maintained in a replicon, such as an extrachromosomal element (eg. plasmids) capable of stable maintenance in a host, such as yeast or bacteria. The replicon may have two replication systems, thus allowing it to be maintained, for example, in yeast for expression and in a prokaryotic host for cloning and amplification. Examples of such yeast-bacteria shuttle vectors include YEp24 [Botstein *et al.* (1979) *Gene* 8:17-24], pCI/1 [Brake *et al.* (1984) *Proc. Natl. Acad. Sci. USA* 81:4642-4646], and YRp17 [Stinchcomb *et al.* (1982) *J. Mol. Biol.* 158:157]. In addition, a replicon may be either a high or low copy number plasmid. A high copy number plasmid will generally have a copy number ranging from about 5 to about 200, and usually about 10 to about 150. A host containing a high copy number plasmid will preferably have at least about 10, and more preferably at least about 20. Enter a high or low copy number vector may be selected, depending upon the effect of the vector and the foreign protein on the host. See eg. Brake *et al.*, *supra*.

Alternatively, the expression constructs can be integrated into the yeast genome with an integrating vector. Integrating vectors usually contain at least one sequence homologous to a yeast chromosome that allows the vector to integrate, and preferably contain two homologous sequences flanking the expression construct. Integrations appear to result from recombinations between homologous DNA in the vector and the yeast chromosome [Orr-Weaver *et al.* (1983) *Methods in Enzymol.* 101:228-245]. An integrating vector may be directed to a specific locus in yeast by selecting the appropriate homologous sequence for inclusion in the vector. See Orr-Weaver *et al.*, *supra*. One or more expression construct may integrate, possibly affecting levels of recombinant protein produced [Rine *et al.* (1983) *Proc. Natl. Acad. Sci. USA* 80:6750]. The chromosomal sequences included in the vector can occur either as a single segment in the vector, which results in the integration of the entire vector, or two segments homologous to adjacent segments in the chromosome and flanking the expression construct in the vector, which can result in the stable integration of only the expression construct.

Usually, extrachromosomal and integrating expression constructs may contain selectable markers to allow for the selection of yeast strains that have been transformed. Selectable markers may include biosynthetic genes that can be expressed in the yeast host, such as *ADE2*, *HIS4*, *LEU2*,

TRP1, and *ALG7*, and the G418 resistance gene, which confer resistance in yeast cells to tunicamycin and G418, respectively. In addition, a suitable selectable marker may also provide yeast with the ability to grow in the presence of toxic compounds, such as metal. For example, the presence of *CUP1* allows yeast to grow in the presence of copper ions [Butt *et al.* (1987) *Microbiol. Rev.* 51:351].

Alternatively, some of the above described components can be put together into transformation vectors. Transformation vectors are usually comprised of a selectable marker that is either maintained in a replicon or developed into an integrating vector, as described above.

Expression and transformation vectors, either extrachromosomal replicons or integrating vectors, have been developed for transformation into many yeasts. For example, expression vectors have been developed for, *inter alia*, the following yeasts: *Candida albicans* [Kurtz, *et al.* (1986) *Mol. Cell. Biol.* 6:142], *Candida maltosa* [Kunze, *et al.* (1985) *J. Basic Microbiol.* 25:141], *Hansenula polymorpha* [Gleeson, *et al.* (1986) *J. Gen. Microbiol.* 132:3459; Roggenkamp *et al.* (1986) *Mol. Gen. Genet.* 202:302], *Kluyveromyces fragilis* [Das, *et al.* (1984) *J. Bacteriol.* 158:1165], *Kluyveromyces lactis* [De Louvencourt *et al.* (1983) *J. Bacteriol.* 154:737; Van den Berg *et al.* (1990) *Bio/Technology* 8:135], *Pichia guilliermondii* [Kunze *et al.* (1985) *J. Basic Microbiol.* 25:141], *Pichia pastoris* [Cregg, *et al.* (1985) *Mol. Cell. Biol.* 5:3376; US Patent Nos. 4,837,148 and 4,929,555], *Saccharomyces cerevisiae* [Hinnen *et al.* (1978) *Proc. Natl. Acad. Sci. USA* 75:1929; Ito *et al.* (1983) *J. Bacteriol.* 153:163], *Schizosaccharomyces pombe* [Beach and Nurse (1981) *Nature* 300:706], and *Yarrowia lipolytica* [Davidow, *et al.* (1985) *Curr. Genet.* 10:380471 Gaillardin, *et al.* (1985) *Curr. Genet.* 10:49].

Methods of introducing exogenous DNA into yeast hosts are well-known in the art, and usually include either the transformation of spheroplasts or of intact yeast cells treated with alkali cations. Transformation procedures usually vary with the yeast species to be transformed. See *eg.* [Kurtz *et al.* (1986) *Mol. Cell. Biol.* 6:142; Kunze *et al.* (1985) *J. Basic Microbiol.* 25:141; *Candida*; [Gleeson *et al.* (1986) *J. Gen. Microbiol.* 132:3459; Roggenkamp *et al.* (1986) *Mol. Gen. Genet.* 202:302; *Hansenula*]; [Das *et al.* (1984) *J. Bacteriol.* 158:1165; De Louvencourt *et al.* (1983) *J. Bacteriol.* 154:1165; Van den Berg *et al.* (1990) *Bio/Technology* 8:135; *Kluyveromyces*]; [Cregg *et al.* (1985) *Mol. Cell. Biol.* 5:3376; Kunze *et al.* (1985) *J. Basic Microbiol.* 25:141; US Patent Nos. 4,837,148 and 4,929,555; *Pichia*]; [Hinnen *et al.* (1978) *Proc. Natl. Acad. Sci. USA* 75:1929;

Ito *et al.* (1983) *J. Bacteriol.* 153:163 Saccharomyces]; [Beach and Nurse (1981) *Nature* 300:706; Schizosaccharomyces]; [Davidow *et al.* (1985) *Curr. Genet.* 10:39; Gaillardin *et al.* (1985) *Curr. Genet.* 10:49; Yarrowia].

Antibodies

- 5 As used herein, the term "antibody" refers to a polypeptide or group of polypeptides composed of at least one antibody combining site. An "antibody combining site" is the three-dimensional binding space with an internal surface shape and charge distribution complementary to the features of an epitope of an antigen, which allows a binding of the antibody with the antigen. "Antibody" includes, for example, vertebrate antibodies, hybrid antibodies, chimeric antibodies, humanised
- 10 antibodies, altered antibodies, univalent antibodies, Fab proteins, and single domain antibodies.

Antibodies against the proteins of the invention are useful for affinity chromatography, immunoassays, and distinguishing/identifying Neisserial proteins.

- Antibodies to the proteins of the invention, both polyclonal and monoclonal, may be prepared by conventional methods. In general, the protein is first used to immunize a suitable animal, preferably
- 15 a mouse, rat, rabbit or goat. Rabbits and goats are preferred for the preparation of polyclonal sera due to the volume of serum obtainable, and the availability of labeled anti-rabbit and anti-goat antibodies. Immunization is generally performed by mixing or emulsifying the protein in saline, preferably in an adjuvant such as Freund's complete adjuvant, and injecting the mixture or emulsion parenterally (generally subcutaneously or intramuscularly). A dose of 50-200 µg/injection
- 20 is typically sufficient. Immunization is generally boosted 2-6 weeks later with one or more injections of the protein in saline, preferably using Freund's incomplete adjuvant. One may alternatively generate antibodies by *in vitro* immunization using methods known in the art, which for the purposes of this invention is considered equivalent to *in vivo* immunization. Polyclonal antisera is obtained by bleeding the immunized animal into a glass or plastic container, incubating
- 25 the blood at 25°C for one hour, followed by incubating at 4°C for 2-18 hours. The serum is recovered by centrifugation (eg. 1,000g for 10 minutes). About 20-50 ml per bleed may be obtained from rabbits.

Monoclonal antibodies are prepared using the standard method of Kohler & Milstein [*Nature* (1975) 256:495-96], or a modification thereof. Typically, a mouse or rat is immunized as described

above. However, rather than bleeding the animal to extract serum, the spleen (and optionally several large lymph nodes) is removed and dissociated into single cells. If desired, the spleen cells may be screened (after removal of nonspecifically adherent cells) by applying a cell suspension to a plate or well coated with the protein antigen. B-cells expressing membrane-bound immunoglobulin specific for the antigen bind to the plate, and are not rinsed away with the rest of the suspension. Resulting B-cells, or all dissociated spleen cells, are then induced to fuse with myeloma cells to form hybridomas, and are cultured in a selective medium (eg. hypoxanthine, aminopterin, thymidine medium, "HAT"). The resulting hybridomas are plated by limiting dilution, and are assayed for the production of antibodies which bind specifically to the immunizing antigen (and which do not bind to unrelated antigens). The selected MAb-secreting hybridomas are then cultured either *in vitro* (eg. in tissue culture bottles or hollow fiber reactors), or *in vivo* (as ascites in mice).

If desired, the antibodies (whether polyclonal or monoclonal) may be labeled using conventional techniques. Suitable labels include fluorophores, chromophores, radioactive atoms (particularly ^{32}P and ^{125}I), electron-dense reagents, enzymes, and ligands having specific binding partners. Enzymes are typically detected by their activity. For example, horseradish peroxidase is usually detected by its ability to convert 3,3',5,5'-tetramethylbenzidine (TMB) to a blue pigment, quantifiable with a spectrophotometer. "Specific binding partner" refers to a protein capable of binding a ligand molecule with high specificity, as for example in the case of an antigen and a monoclonal antibody specific therefor. Other specific binding partners include biotin and avidin or streptavidin, IgG and protein A, and the numerous receptor-ligand couples known in the art. It should be understood that the above description is not meant to categorize the various labels into distinct classes, as the same label may serve in several different modes. For example, ^{125}I may serve as a radioactive label or as an electron-dense reagent. HRP may serve as enzyme or as antigen for a MAb. Further, one may combine various labels for desired effect. For example, MAbs and avidin also require labels in the practice of this invention: thus, one might label a MAb with biotin, and detect its presence with avidin labeled with ^{125}I , or with an anti-biotin MAb labeled with HRP. Other permutations and possibilities will be readily apparent to those of ordinary skill in the art, and are considered as equivalents within the scope of the instant invention.

Pharmaceutical Compositions

Pharmaceutical compositions can comprise either polypeptides, antibodies, or nucleic acid of the invention. The pharmaceutical compositions will comprise a therapeutically effective amount of either polypeptides, antibodies, or polynucleotides of the claimed invention.

- 5 The term "therapeutically effective amount" as used herein refers to an amount of a therapeutic agent to treat, ameliorate, or prevent a desired disease or condition, or to exhibit a detectable therapeutic or preventative effect. The effect can be detected by, for example, chemical markers or antigen levels. Therapeutic effects also include reduction in physical symptoms, such as decreased body temperature. The precise effective amount for a subject will depend upon the subject's size
10 and health, the nature and extent of the condition, and the therapeutics or combination of therapeutics selected for administration. Thus, it is not useful to specify an exact effective amount in advance. However, the effective amount for a given situation can be determined by routine experimentation and is within the judgement of the clinician.

- For purposes of the present invention, an effective dose will be from about 0.01 mg/kg to 50 mg/kg
15 or 0.05 mg/kg to about 10 mg/kg of the DNA constructs in the individual to which it is administered.

- A pharmaceutical composition can also contain a pharmaceutically acceptable carrier. The term "pharmaceutically acceptable carrier" refers to a carrier for administration of a therapeutic agent, such as antibodies or a polypeptide, genes, and other therapeutic agents. The term refers to any
20 pharmaceutical carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition, and which may be administered without undue toxicity. Suitable carriers may be large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, and inactive virus particles. Such carriers are well known to those of ordinary skill in
25 the art.

Pharmaceutically acceptable salts can be used therein, for example, mineral acid salts such as hydrochlorides, hydrobromides, phosphates, sulfates, and the like; and the salts of organic acids such as acetates, propionates, malonates, benzoates, and the like. A thorough discussion of

pharmaceutically acceptable excipients is available in Remington's Pharmaceutical Sciences (Mack Pub. Co., N.J. 1991).

Pharmaceutically acceptable carriers in therapeutic compositions may contain liquids such as water, saline, glycerol and ethanol. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such vehicles. Typically, the therapeutic compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection may also be prepared. Liposomes are included within the definition of a pharmaceutically acceptable carrier.

Delivery Methods

Once formulated, the compositions of the invention can be administered directly to the subject. The subjects to be treated can be animals; in particular, human subjects can be treated.

Direct delivery of the compositions will generally be accomplished by injection, either subcutaneously, intraperitoneally, intravenously or intramuscularly or delivered to the interstitial space of a tissue. The compositions can also be administered into a lesion. Other modes of administration include oral and pulmonary administration, suppositories, and transdermal or transcutaneous applications (eg. see WO98/20734), needles, and gene guns or hyposprays. Dosage treatment may be a single dose schedule or a multiple dose schedule.

Vaccines

Vaccines according to the invention may either be prophylactic (ie. to prevent infection) or therapeutic (ie. to treat disease after infection).

Such vaccines comprise immunising antigen(s), immunogen(s), polypeptide(s), protein(s) or nucleic acid, usually in combination with "pharmaceutically acceptable carriers," which include any carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition. Suitable carriers are typically large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, lipid aggregates (such as oil droplets or liposomes), and inactive virus particles. Such carriers are well known to those of ordinary skill in the art. Additionally, these carriers may function as immunostimulating agents ("adjuvants"). Furthermore, the antigen or immunogen may be conjugated to a bacterial toxoid, such as a toxoid from diphtheria, tetanus, cholera, *H. pylori*, etc. pathogens.

Preferred adjuvants to enhance effectiveness of the composition include, but are not limited to: (1) aluminum salts (alum), such as aluminum hydroxide, aluminum phosphate, aluminum sulfate, etc; (2) oil-in-water emulsion formulations (with or without other specific immunostimulating agents such as muramyl peptides (see below) or bacterial cell wall components), such as for example (a) MF59™ (WO 90/14837; Chapter 10 in *Vaccine design: the subunit and adjuvant approach*, eds. Powell & Newman, Plenum Press 1995), containing 5% Squalene, 0.5% Tween 80, and 0.5% Span 85 (optionally containing various amounts of MTP-PE (see below), although not required) formulated into submicron particles using a microfluidizer such as Model 110Y microfluidizer (Microfluidics, Newton, MA), (b) SAF, containing 10% Squalene, 0.4% Tween 80, 5% pluronic-blocked polymer L121, and thr-MDP (see below) either microfluidized into a submicron emulsion or vortexed to generate a larger particle size emulsion, and (c) Ribi™ adjuvant system (RAS), (Ribi Immunochem, Hamilton, MT) containing 2% Squalene, 0.2% Tween 80, and one or more bacterial cell wall components from the group consisting of monophosphorylipid A (MPL), trehalose dimycolate (TDM), and cell wall skeleton (CWS), preferably MPL + CWS (Detox™); (3) saponin adjuvants, such as Stimulon™ (Cambridge Bioscience, Worcester, MA) may be used or particles generated therefrom such as ISCOMs (immunostimulating complexes); (4) Complete Freund's Adjuvant (CFA) and Incomplete Freund's Adjuvant (IFA); (5) cytokines, such as interleukins (eg. IL-1, IL-2, IL-4, IL-5, IL-6, IL-7, IL-12, etc.), interferons (eg. gamma interferon), macrophage colony stimulating factor (M-CSF), tumor necrosis factor (TNF), etc; and (6) other substances that act as immunostimulating agents to enhance the effectiveness of the composition. Alum and MF59™ are preferred.

As mentioned above, muramyl peptides include, but are not limited to, N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP), N-acetyl-normuramyl-L-alanyl-D-isoglutamine (nor-MDP), N-acetylmuramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl-sn-glycero-3-hydroxyphosphoryloxy)-ethylamine (MTP-PE), etc.

The immunogenic compositions (eg. the immunising antigen/immunogen/polypeptide/protein/nucleic acid, pharmaceutically acceptable carrier, and adjuvant) typically will contain diluents, such as water, saline, glycerol, ethanol, etc. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such vehicles.

Typically, the immunogenic compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection may also be prepared. The preparation also may be emulsified or encapsulated in liposomes for enhanced adjuvant effect, as discussed above under pharmaceutically acceptable carriers.

- 5 Immunogenic compositions used as vaccines comprise an immunologically effective amount of the antigenic or immunogenic polypeptides, as well as any other of the above-mentioned components, as needed. By "immunologically effective amount", it is meant that the administration of that amount to an individual, either in a single dose or as part of a series, is effective for treatment or prevention. This amount varies depending upon the health and physical condition of the individual
- 10 to be treated, the taxonomic group of individual to be treated (*eg.* nonhuman primate, primate, *etc.*), the capacity of the individual's immune system to synthesize antibodies, the degree of protection desired, the formulation of the vaccine, the treating doctor's assessment of the medical situation, and other relevant factors. It is expected that the amount will fall in a relatively broad range that can be determined through routine trials.
- 15 The immunogenic compositions are conventionally administered parenterally, *eg.* by injection, either subcutaneously, intramuscularly, or transdermally/transcutaneously (*eg.* WO98/20734). Additional formulations suitable for other modes of administration include oral and pulmonary formulations, suppositories, and transdermal applications. Dosage treatment may be a single dose schedule or a multiple dose schedule. The vaccine may be administered in conjunction with other
- 20 immunoregulatory agents.

As an alternative to protein-based vaccines, DNA vaccination may be employed [*eg.* Robinson & Torres (1997) *Seminars in Immunology* 9:271-283; Donnelly *et al.* (1997) *Annu Rev Immunol* 15:617-648; see later herein].

Gene Delivery Vehicles

- 25 Gene therapy vehicles for delivery of constructs including a coding sequence of a therapeutic of the invention, to be delivered to the mammal for expression in the mammal, can be administered either locally or systemically. These constructs can utilize viral or non-viral vector approaches in *in vivo* or *ex vivo* modality. Expression of such coding sequence can be induced using endogenous

mammalian or heterologous promoters. Expression of the coding sequence in vivo can be either constitutive or regulated.

The invention includes gene delivery vehicles capable of expressing the contemplated nucleic acid sequences. The gene delivery vehicle is preferably a viral vector and, more preferably, a retroviral, adenoviral, adeno-associated viral (AAV), herpes viral, or alphavirus vector. The viral vector can also be an astrovirus, coronavirus, orthomyxovirus, papovavirus, paramyxovirus, parvovirus, picornavirus, poxvirus, or togavirus viral vector. See generally, Jolly (1994) *Cancer Gene Therapy* 1:51-64; Kimura (1994) *Human Gene Therapy* 5:845-852; Connelly (1995) *Human Gene Therapy* 6:185-193; and Kaplitt (1994) *Nature Genetics* 6:148-153.

Retroviral vectors are well known in the art and we contemplate that any retroviral gene therapy vector is employable in the invention, including B, C and D type retroviruses, xenotropic retroviruses (for example, NZB-X1, NZB-X2 and NZB9-1 (see O'Neill (1985) *J. Virol.* 53:160) polytropic retroviruses eg. MCF and MCF-MLV (see Kelly (1983) *J. Virol.* 45:291), spumaviruses and lentiviruses. See RNA Tumor Viruses, Second Edition, Cold Spring Harbor Laboratory, 1985.

Portions of the retroviral gene therapy vector may be derived from different retroviruses. For example, retrovector LTRs may be derived from a Murine Sarcoma Virus, a tRNA binding site from a Rous Sarcoma Virus, a packaging signal from a Murine Leukemia Virus, and an origin of second strand synthesis from an Avian Leukosis Virus.

These recombinant retroviral vectors may be used to generate transduction competent retroviral vector particles by introducing them into appropriate packaging cell lines (see US patent 5,591,624). Retrovirus vectors can be constructed for site-specific integration into host cell DNA by incorporation of a chimeric integrase enzyme into the retroviral particle (see WO96/37626). It is preferable that the recombinant viral vector is a replication defective recombinant virus.

Packaging cell lines suitable for use with the above-described retrovirus vectors are well known in the art, are readily prepared (see WO95/30763 and WO92/05266), and can be used to create producer cell lines (also termed vector cell lines or "VCLs") for the production of recombinant vector particles. Preferably, the packaging cell lines are made from human parent cells (eg. HT1080 cells) or mink parent cell lines, which eliminates inactivation in human serum.

Preferred retroviruses for the construction of retroviral gene therapy vectors include Avian Leukosis Virus, Bovine Leukemia Virus, Murine Leukemia Virus, Mink-Cell Focus-Inducing Virus, Murine Sarcoma Virus, Reticuloendotheliosis Virus and Rous Sarcoma Virus. Particularly preferred Murine Leukemia Viruses include 4070A and 1504A (Hartley and Rowe (1976) *J Virol* 19:19-25), Abelson (ATCC No. VR-999), Friend (ATCC No. VR-245), Graffi, Gross (ATCC No. VR-590), Kirsten, Harvey Sarcoma Virus and Rauscher (ATCC No. VR-998) and Moloney Murine Leukemia Virus (ATCC No. VR-190). Such retroviruses may be obtained from depositories or collections such as the American Type Culture Collection ("ATCC") in Rockville, Maryland or isolated from known sources using commonly available techniques.

Exemplary known retroviral gene therapy vectors employable in this invention include those described in patent applications GB2200651, EP0415731, EP0345242, EP0334301, WO89/02468; WO89/05349, WO89/09271, WO90/02806, WO90/07936, WO94/03622, WO93/25698, WO93/25234, WO93/11230, WO93/10218, WO91/02805, WO91/02825, WO95/07994, US 5,219,740, US 4,405,712, US 4,861,719, US 4,980,289, US 4,777,127, US 5,591,624. See also Vile (1993) *Cancer Res* 53:3860-3864; Vile (1993) *Cancer Res* 53:962-967; Ram (1993) *Cancer Res* 53 (1993) 83-88; Takamiya (1992) *J Neurosci Res* 33:493-503; Baba (1993) *J Neurosurg* 79:729-735; Mann (1983) *Cell* 33:153; Cane (1984) *Proc Natl Acad Sci* 81:6349; and Miller (1990) *Human Gene Therapy* 1.

Human adenoviral gene therapy vectors are also known in the art and employable in this invention.

See, for example, Berkner (1988) *Biotechniques* 6:616 and Rosenfeld (1991) *Science* 252:431, and WO93/07283, WO93/06223, and WO93/07282. Exemplary known adenoviral gene therapy vectors employable in this invention include those described in the above referenced documents and in WO94/12649, WO93/03769, WO93/19191, WO94/28938, WO95/11984, WO95/00655, WO95/27071, WO95/29993, WO95/34671, WO96/05320, WO94/08026, WO94/11506, WO93/06223, WO94/24299, WO95/14102, WO95/24297, WO95/02697, WO94/28152, WO94/24299, WO95/09241, WO95/25807, WO95/05835, WO94/18922 and WO95/09654. Alternatively, administration of DNA linked to killed adenovirus as described in Curiel (1992) *Hum. Gene Ther.* 3:147-154 may be employed. The gene delivery vehicles of the invention also include adenovirus associated virus (AAV) vectors. Leading and preferred examples of such vectors for use in this invention are the AAV-2 based vectors disclosed in Srivastava,

WO93/09239. Most preferred AAV vectors comprise the two AAV inverted terminal repeats in which the native D-sequences are modified by substitution of nucleotides, such that at least 5 native nucleotides and up to 18 native nucleotides, preferably at least 10 native nucleotides up to 18 native nucleotides, most preferably 10 native nucleotides are retained and the remaining nucleotides of the D-sequence are deleted or replaced with non-native nucleotides. The native D-sequences of the AAV inverted terminal repeats are sequences of 20 consecutive nucleotides in each AAV inverted terminal repeat (*ie.* there is one sequence at each end) which are not involved in HP formation. The non-native replacement nucleotide may be any nucleotide other than the nucleotide found in the native D-sequence in the same position. Other employable exemplary AAV vectors are pWP-19, pWN-1, both of which are disclosed in Nahreini (1993) *Gene* 124:257-262. Another example of such an AAV vector is psub201 (see Samulski (1987) *J. Virol.* 61:3096). Another exemplary AAV vector is the Double-D ITR vector. Construction of the Double-D ITR vector is disclosed in US Patent 5,478,745. Still other vectors are those disclosed in Carter US Patent 4,797,368 and Muzyczka US Patent 5,139,941, Chartejee US Patent 5,474,935, and Kotin WO94/288157. Yet a further example of an AAV vector employable in this invention is SSV9AFABTKneo, which contains the AFP enhancer and albumin promoter and directs expression predominantly in the liver. Its structure and construction are disclosed in Su (1996) *Human Gene Therapy* 7:463-470. Additional AAV gene therapy vectors are described in US 5,354,678, US 5,173,414, US 5,139,941, and US 5,252,479.

The gene therapy vectors of the invention also include herpes vectors. Leading and preferred examples are herpes simplex virus vectors containing a sequence encoding a thymidine kinase polypeptide such as those disclosed in US 5,288,641 and EP0176170 (Roizman). Additional exemplary herpes simplex virus vectors include HFEM/ICP6-LacZ disclosed in WO95/04139 (Wistar Institute), pHSVlac described in Geller (1988) *Science* 241:1667-1669 and in WO90/09441 and WO92/07945, HSV Us3::pgC-lacZ described in Fink (1992) *Human Gene Therapy* 3:11-19 and HSV 7134, 2 RH 105 and GAL4 described in EP 0453242 (Breakefield), and those deposited with the ATCC as accession numbers ATCC VR-977 and ATCC VR-260.

Also contemplated are alpha virus gene therapy vectors that can be employed in this invention. Preferred alpha virus vectors are Sindbis viruses vectors. Togaviruses, Semliki Forest virus (ATCC VR-67; ATCC VR-1247), Middleberg virus (ATCC VR-370), Ross River virus (ATCC VR-373;

ATCC VR-1246), Venezuelan equine encephalitis virus (ATCC VR923; ATCC VR-1250; ATCC VR-1249; ATCC VR-532), and those described in US patents 5,091,309, 5,217,879, and WO92/10578. More particularly, those alpha virus vectors described in US Serial No. 08/405,627, filed March 15, 1995, WO94/21792, WO92/10578, WO95/07994, US 5,091,309 and US 5,217,879 are employable. Such alpha viruses may be obtained from depositories or collections such as the ATCC in Rockville, Maryland or isolated from known sources using commonly available techniques. Preferably, alphavirus vectors with reduced cytotoxicity are used (see USSN 08/679640).

DNA vector systems such as eukaryotic layered expression systems are also useful for expressing the nucleic acids of the invention. See WO95/07994 for a detailed description of eukaryotic layered expression systems. Preferably, the eukaryotic layered expression systems of the invention are derived from alphavirus vectors and most preferably from Sindbis viral vectors.

Other viral vectors suitable for use in the present invention include those derived from poliovirus, for example ATCC VR-58 and those described in Evans, Nature 339 (1989) 385 and Sabin (1973)

J. Biol. Standardization 1:115; rhinovirus, for example ATCC VR-1110 and those described in Arnold (1990) *J Cell Biochem* L401; pox viruses such as canary pox virus or vaccinia virus, for example ATCC VR-111 and ATCC VR-2010 and those described in Fisher-Hoch (1989) *Proc Natl Acad Sci* 86:317; Flexner (1989) *Ann NY Acad Sci* 569:86, Flexner (1990) *Vaccine* 8:17; in US 4,603,112 and US 4,769,330 and WO89/01973; SV40 virus, for example ATCC VR-305 and those

described in Mulligan (1979) *Nature* 277:108 and Madzak (1992) *J Gen Virol* 73:1533; influenza virus, for example ATCC VR-797 and recombinant influenza viruses made employing reverse genetics techniques as described in US 5,166,057 and in Enami (1990) *Proc Natl Acad Sci* 87:3802-3805; Enami & Palese (1991) *J Virol* 65:2711-2713 and Luytjes (1989) *Cell* 59:110, (see also McMichael (1983) *NEJ Med* 309:13, and Yap (1978) *Nature* 273:238 and *Nature* (1979)

277:108); human immunodeficiency virus as described in EP-0386882 and in Buchschacher (1992) *J. Virol.* 66:2731; measles virus, for example ATCC VR-67 and VR-1247 and those described in EP-0440219; Aura virus, for example ATCC VR-368; Bebaru virus, for example ATCC VR-600 and ATCC VR-1240; Cabassou virus, for example ATCC VR-922; Chikungunya virus, for example ATCC VR-64 and ATCC VR-1241; Fort Morgan Virus, for example ATCC VR-924;

Getah virus, for example ATCC VR-369 and ATCC VR-1243; Kyzylagach virus, for example

ATCC VR-927; Mayaro virus, for example ATCC VR-66; Mucambo virus, for example ATCC VR-580 and ATCC VR-1244; Ndumu virus, for example ATCC VR-371; Pixuna virus, for example ATCC VR-372 and ATCC VR-1245; Tonate virus, for example ATCC VR-925; Trinit virus, for example ATCC VR-469; Una virus, for example ATCC VR-374; Whataroa virus, for example ATCC VR-926; Y-62-33 virus, for example ATCC VR-375; O'Nyong virus, Eastern encephalitis virus, for example ATCC VR-65 and ATCC VR-1242; Western encephalitis virus, for example ATCC VR-70, ATCC VR-1251, ATCC VR-622 and ATCC VR-1252; and coronavirus, for example ATCC VR-740 and those described in Hamre (1966) *Proc Soc Exp Biol Med* 121:190.

Delivery of the compositions of this invention into cells is not limited to the above mentioned viral vectors. Other delivery methods and media may be employed such as, for example, nucleic acid expression vectors, polycationic condensed DNA linked or unlinked to killed adenovirus alone, for example see US Serial No. 08/366,787, filed December 30, 1994 and Curiel (1992) *Hum Gene Ther* 3:147-154 ligand linked DNA, for example see Wu (1989) *J Biol Chem* 264:16985-16987, eucaryotic cell delivery vehicles cells, for example see US Serial No.08/240,030, filed May 9, 1994, and US Serial No. 08/404,796, deposition of photopolymerized hydrogel materials, hand-held gene transfer particle gun, as described in US Patent 5,149,655, ionizing radiation as described in US5,206,152 and in WO92/11033, nucleic charge neutralization or fusion with cell membranes. Additional approaches are described in Philip (1994) *Mol Cell Biol* 14:2411-2418 and in Woffendin (1994) *Proc Natl Acad Sci* 91:1581-1585.

Particle mediated gene transfer may be employed, for example see US Serial No. 60/023,867. Briefly, the sequence can be inserted into conventional vectors that contain conventional control sequences for high level expression, and then incubated with synthetic gene transfer molecules such as polymeric DNA-binding cations like polylysine, protamine, and albumin, linked to cell targeting ligands such as asialoorosomucoid, as described in Wu & Wu (1987) *J. Biol. Chem.* 262:4429-4432, insulin as described in Hucked (1990) *Biochem Pharmacol* 40:253-263, galactose as described in Plank (1992) *Bioconjugate Chem* 3:533-539, lactose or transferrin.

Naked DNA may also be employed. Exemplary naked DNA introduction methods are described in WO 90/11092 and US 5,580,859. Uptake efficiency may be improved using biodegradable latex beads. DNA coated latex beads are efficiently transported into cells after endocytosis initiation by the beads. The method may be improved further by treatment of the beads to increase

hydrophobicity and thereby facilitate disruption of the endosome and release of the DNA into the cytoplasm.

Liposomes that can act as gene delivery vehicles are described in US 5,422,120, WO95/13796, WO94/23697, WO91/14445 and EP-524,968. As described in USSN. 60/023,867, on non-viral delivery, the nucleic acid sequences encoding a polypeptide can be inserted into conventional vectors that contain conventional control sequences for high level expression, and then be incubated with synthetic gene transfer molecules such as polymeric DNA-binding cations like polylysine, protamine, and albumin, linked to cell targeting ligands such as asialoorosomucoid, insulin, galactose, lactose, or transferrin. Other delivery systems include the use of liposomes to encapsulate DNA comprising the gene under the control of a variety of tissue-specific or ubiquitously-active promoters. Further non-viral delivery suitable for use includes mechanical delivery systems such as the approach described in Woffendin *et al* (1994) *Proc. Natl. Acad. Sci. USA* 91(24):11581-11585. Moreover, the coding sequence and the product of expression of such can be delivered through deposition of photopolymerized hydrogel materials. Other conventional methods for gene delivery that can be used for delivery of the coding sequence include, for example, use of hand-held gene transfer particle gun, as described in US 5,149,655; use of ionizing radiation for activating transferred gene, as described in US 5,206,152 and WO92/11033

Exemplary liposome and polycationic gene delivery vehicles are those described in US 5,422,120 and 4,762,915; in WO 95/13796; WO94/23697; and WO91/14445; in EP-0524968; and in Stryer, Biochemistry, pages 236-240 (1975) W.H. Freeman, San Francisco; Szoka (1980) *Biochem Biophys Acta* 600:1; Bayer (1979) *Biochem Biophys Acta* 550:464; Rivnay (1987) *Meth Enzymol* 149:119; Wang (1987) *Proc Natl Acad Sci* 84:7851; Plant (1989) *Anal Biochem* 176:420.

A polynucleotide composition can comprises therapeutically effective amount of a gene therapy vehicle, as the term is defined above. For purposes of the present invention, an effective dose will be from about 0.01 mg/ kg to 50 mg/kg or 0.05 mg/kg to about 10 mg/kg of the DNA constructs in the individual to which it is administered.

Delivery Methods

Once formulated, the polynucleotide compositions of the invention can be administered (1) directly to the subject; (2) delivered *ex vivo*, to cells derived from the subject; or (3) *in vitro* for expression

of recombinant proteins. The subjects to be treated can be mammals or birds. Also, human subjects can be treated.

Direct delivery of the compositions will generally be accomplished by injection, either subcutaneously, intraperitoneally, intravenously or intramuscularly or delivered to the interstitial space of a tissue. The compositions can also be administered into a lesion. Other modes of administration include oral and pulmonary administration, suppositories, and transdermal or transcutaneous applications (eg. see WO98/20734), needles, and gene guns or hyposprays. Dosage treatment may be a single dose schedule or a multiple dose schedule.

Methods for the *ex vivo* delivery and reimplantation of transformed cells into a subject are known in the art and described in eg. WO93/14778. Examples of cells useful in *ex vivo* applications include, for example, stem cells, particularly hematopoietic, lymph cells, macrophages, dendritic cells, or tumor cells.

Generally, delivery of nucleic acids for both *ex vivo* and *in vitro* applications can be accomplished by the following procedures, for example, dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the DNA into nuclei, all well known in the art.

Polynucleotide and polypeptide pharmaceutical compositions

In addition to the pharmaceutically acceptable carriers and salts described above, the following additional agents can be used with polynucleotide and/or polypeptide compositions.

A. Polypeptides

One example are polypeptides which include, without limitation: asioloorosomucoid (ASOR); transferrin; asialoglycoproteins; antibodies; antibody fragments; ferritin; interleukins; interferons, granulocyte, macrophage colony stimulating factor (GM-CSF), granulocyte colony stimulating factor (G-CSF), macrophage colony stimulating factor (M-CSF), stem cell factor and erythropoietin. Viral antigens, such as envelope proteins, can also be used. Also, proteins from other invasive organisms, such as the 17 amino acid peptide from the circumsporozoite protein of plasmodium falciparum known as RII.

B.Hormones, Vitamins, etc.

Other groups that can be included are, for example: hormones, steroids, androgens, estrogens, thyroid hormone, or vitamins, folic acid.

C.Polyalkylenes, Polysaccharides, etc.

- 5 Also, polyalkylene glycol can be included with the desired polynucleotides/polypeptides. In a preferred embodiment, the polyalkylene glycol is polyethylene glycol. In addition, mono-, di-, or polysaccharides can be included. In a preferred embodiment of this aspect, the polysaccharide is dextran or DEAE-dextran. Also, chitosan and poly(lactide-co-glycolide)

D.Lipids, and Liposomes

- 10 The desired polynucleotide/polypeptide can also be encapsulated in lipids or packaged in liposomes prior to delivery to the subject or to cells derived therefrom.

Lipid encapsulation is generally accomplished using liposomes which are able to stably bind or entrap and retain nucleic acid. The ratio of condensed polynucleotide to lipid preparation can vary but will generally be around 1:1 (mg DNA:micromoles lipid), or more of lipid. For a review of the

- 15 use of liposomes as carriers for delivery of nucleic acids, see, Hug and Sleight (1991) *Biochim. Biophys. Acta*. 1097:1-17; Straubinger (1983) *Meth. Enzymol.* 101:512-527.

Liposomal preparations for use in the present invention include cationic (positively charged), anionic (negatively charged) and neutral preparations. Cationic liposomes have been shown to mediate intracellular delivery of plasmid DNA (Felgner (1987) *Proc. Natl. Acad. Sci. USA* 84:7413-7416); mRNA (Malone (1989) *Proc. Natl. Acad. Sci. USA* 86:6077-6081); and purified transcription factors (Debs (1990) *J. Biol. Chem.* 265:10189-10192), in functional form.

- 20 Cationic liposomes are readily available. For example, N[1-2,3-dioleoyloxy]propyl]-N,N,N-triethylammonium (DOTMA) liposomes are available under the trademark Lipofectin, from GIBCO BRL, Grand Island, NY. (See, also, Felgner *supra*). Other commercially available liposomes include transfectace (DDAB/DOPE) and DOTAP/DOPE (Boehringer). Other cationic liposomes can be prepared from readily available materials using techniques well known in the art. See, eg. Szoka (1978) *Proc. Natl. Acad. Sci. USA* 75:4194-4198; WO90/11092 for a description of the synthesis of DOTAP (1,2-bis(oleoyloxy)-3-(trimethylammonio)propane) liposomes.

Similarly, anionic and neutral liposomes are readily available, such as from Avanti Polar Lipids (Birmingham, AL), or can be easily prepared using readily available materials. Such materials include phosphatidyl choline, cholesterol, phosphatidyl ethanolamine, dioleoylphosphatidyl choline (DOPC), dioleoylphosphatidyl glycerol (DOPG), dioleoylphosphatidyl ethanolamine (DOPE),
5 among others. These materials can also be mixed with the DOTMA and DOTAP starting materials in appropriate ratios. Methods for making liposomes using these materials are well known in the art.

The liposomes can comprise multilammellar vesicles (MLVs), small unilamellar vesicles (SUVs), or large unilamellar vesicles (LUVs). The various liposome-nucleic acid complexes are prepared using methods known in the art. See eg. Straubinger (1983) *Meth. Immunol.* 101:512-527; Szoka
10 (1978) *Proc. Natl. Acad. Sci. USA* 75:4194-4198; Papahadjopoulos (1975) *Biochim. Biophys. Acta* 394:483; Wilson (1979) *Cell* 17:77; Deamer & Bangham (1976) *Biochim. Biophys. Acta* 443:629; Ostro (1977) *Biochem. Biophys. Res. Commun.* 76:836; Fraley (1979) *Proc. Natl. Acad. Sci. USA* 76:3348; Enoch & Strittmatter (1979) *Proc. Natl. Acad. Sci. USA* 76:145; Fraley (1980) *J. Biol.*
15 *Chem.* (1980) 255:10431; Szoka & Papahadjopoulos (1978) *Proc. Natl. Acad. Sci. USA* 75:145; and Schaefer-Ridder (1982) *Science* 215:166.

E. Lipoproteins

In addition, lipoproteins can be included with the polynucleotide/polypeptide to be delivered. Examples of lipoproteins to be utilized include: chylomicrons, HDL, IDL, LDL, and VLDL.
20 Mutants, fragments, or fusions of these proteins can also be used. Also, modifications of naturally occurring lipoproteins can be used, such as acetylated LDL. These lipoproteins can target the delivery of polynucleotides to cells expressing lipoprotein receptors. Preferably, if lipoproteins are including with the polynucleotide to be delivered, no other targeting ligand is included in the composition.

25 Naturally occurring lipoproteins comprise a lipid and a protein portion. The protein portion are known as apoproteins. At the present, apoproteins A, B, C, D, and E have been isolated and identified. At least two of these contain several proteins, designated by Roman numerals, AI, AII, AIV; CI, CII, CIII.

A lipoprotein can comprise more than one apoprotein. For example, naturally occurring
30 chylomicrons comprises of A, B, C, and E, over time these lipoproteins lose A and acquire C and

E apoproteins. VLDL comprises A, B, C, and E apoproteins, LDL comprises apoprotein B; and HDL comprises apoproteins A, C, and E.

The amino acid of these apoproteins are known and are described in, for example, Breslow (1985) *Annu Rev. Biochem* 54:699; Law (1986) *Adv. Exp Med. Biol.* 151:162; Chen (1986) *J Biol Chem* 261:12918; Kane (1980) *Proc Natl Acad Sci USA* 77:2465; and Utermann (1984) *Hum Genet* 65:232.

Lipoproteins contain a variety of lipids including, triglycerides, cholesterol (free and esters), and phospholipids. The composition of the lipids varies in naturally occurring lipoproteins. For example, chylomicrons comprise mainly triglycerides. A more detailed description of the lipid content of naturally occurring lipoproteins can be found, for example, in *Meth. Enzymol.* 128 (1986). The composition of the lipids are chosen to aid in conformation of the apoprotein for receptor binding activity. The composition of lipids can also be chosen to facilitate hydrophobic interaction and association with the polynucleotide binding molecule.

Naturally occurring lipoproteins can be isolated from serum by ultracentrifugation, for instance. Such methods are described in *Meth. Enzymol.* (*supra*); Pitas (1980) *J. Biochem.* 255:5454-5460 and Mahey (1979) *J Clin. Invest* 64:743-750. Lipoproteins can also be produced by *in vitro* or recombinant methods by expression of the apoprotein genes in a desired host cell. See, for example, Atkinson (1986) *Annu Rev Biophys Chem* 15:403 and Radding (1958) *Biochim Biophys Acta* 30: 443. Lipoproteins can also be purchased from commercial suppliers, such as Biomedical Technologies, Inc., Stoughton, Massachusetts, USA. Further description of lipoproteins can be found in Zuckermann *et al.* PCT/US97/14465.

F. Polycationic Agents

Polycationic agents can be included, with or without lipoprotein, in a composition with the desired polynucleotide/polypeptide to be delivered.

Polycationic agents, typically, exhibit a net positive charge at physiological relevant pH and are capable of neutralizing the electrical charge of nucleic acids to facilitate delivery to a desired location. These agents have both *in vitro*, *ex vivo*, and *in vivo* applications. Polycationic agents can be used to deliver nucleic acids to a living subject either intramuscularly, subcutaneously, etc.

The following are examples of useful polypeptides as polycationic agents: polylysine, polyarginine, polyornithine, and protamine. Other examples include histones, protamines, human serum albumin, DNA binding proteins, non-histone chromosomal proteins, coat proteins from DNA viruses, such as (X174, transcriptional factors also contain domains that bind DNA and therefore may be useful as nucleic acid condensing agents. Briefly, transcriptional factors such as C/CEBP, c-jun, c-fos, AP-1, AP-2, AP-3, CPF, Prot-1, Sp-1, Oct-1, Oct-2, CREP, and TFIIID contain basic domains that bind DNA sequences.

Organic polycationic agents include: spermine, spermidine, and putrescine.

The dimensions and of the physical properties of a polycationic agent can be extrapolated from the list above, to construct other polypeptide polycationic agents or to produce synthetic polycationic agents.

Synthetic polycationic agents which are useful include, for example, DEAE-dextran, polybrene. Lipofectin™, and lipofectAMINE™ are monomers that form polycationic complexes when combined with polynucleotides/polypeptides.

Immunodiagnostic Assays

Neisserial antigens of the invention can be used in immunoassays to detect antibody levels (or, conversely, anti-Neisserial antibodies can be used to detect antigen levels). Immunoassays based on well defined, recombinant antigens can be developed to replace invasive diagnostics methods. Antibodies to Neisserial proteins within biological samples, including for example, blood or serum samples, can be detected. Design of the immunoassays is subject to a great deal of variation, and a variety of these are known in the art. Protocols for the immunoassay may be based, for example, upon competition, or direct reaction, or sandwich type assays. Protocols may also, for example, use solid supports, or may be by immunoprecipitation. Most assays involve the use of labeled antibody or polypeptide; the labels may be, for example, fluorescent, chemiluminescent, radioactive, or dye molecules. Assays which amplify the signals from the probe are also known; examples of which are assays which utilize biotin and avidin, and enzyme-labeled and mediated immunoassays, such as ELISA assays.

Kits suitable for immunodiagnosis and containing the appropriate labeled reagents are constructed by packaging the appropriate materials, including the compositions of the invention, in suitable

containers, along with the remaining reagents and materials (for example, suitable buffers, salt solutions, *etc.*) required for the conduct of the assay, as well as suitable set of assay instructions.

Nucleic Acid Hybridisation

“Hybridization” refers to the association of two nucleic acid sequences to one another by hydrogen bonding. Typically, one sequence will be fixed to a solid support and the other will be free in solution. Then, the two sequences will be placed in contact with one another under conditions that favor hydrogen bonding. Factors that affect this bonding include: the type and volume of solvent; reaction temperature; time of hybridization; agitation; agents to block the non-specific attachment of the liquid phase sequence to the solid support (Denhardt's reagent or BLOTTO); concentration of the sequences; use of compounds to increase the rate of association of sequences (dextran sulfate or polyethylene glycol); and the stringency of the washing conditions following hybridization. See Sambrook *et al.* [*supra*] Volume 2, chapter 9, pages 9.47 to 9.57.

“Stringency” refers to conditions in a hybridization reaction that favor association of very similar sequences over sequences that differ. For example, the combination of temperature and salt concentration should be chosen that is approximately 120 to 200°C below the calculated T_m of the hybrid under study. The temperature and salt conditions can often be determined empirically in preliminary experiments in which samples of genomic DNA immobilized on filters are hybridized to the sequence of interest and then washed under conditions of different stringencies. See Sambrook *et al.* at page 9.50.

Variables to consider when performing, for example, a Southern blot are (1) the complexity of the DNA being blotted and (2) the homology between the probe and the sequences being detected. The total amount of the fragment(s) to be studied can vary a magnitude of 10, from 0.1 to 1 µg for a plasmid or phage digest to 10^{-9} to 10^{-8} g for a single copy gene in a highly complex eukaryotic genome. For lower complexity polynucleotides, substantially shorter blotting, hybridization, and exposure times, a smaller amount of starting polynucleotides, and lower specific activity of probes can be used. For example, a single-copy yeast gene can be detected with an exposure time of only 1 hour starting with 1 µg of yeast DNA, blotting for two hours, and hybridizing for 4-8 hours with a probe of 10^8 cpm/µg. For a single-copy mammalian gene a conservative approach would start with 10 µg of DNA, blot overnight, and hybridize overnight in the presence of 10% dextran sulfate using a probe of greater than 10^8 cpm/µg, resulting in an exposure time of ~24 hours.

Several factors can affect the melting temperature (T_m) of a DNA-DNA hybrid between the probe and the fragment of interest, and consequently, the appropriate conditions for hybridization and washing. In many cases the probe is not 100% homologous to the fragment. Other commonly encountered variables include the length and total G+C content of the hybridizing sequences and the ionic strength and formamide content of the hybridization buffer. The effects of all of these factors can be approximated by a single equation:

$$T_m = 81 + 16.6(\log_{10} C_i) + 0.4[\%(G + C)] - 0.6(\% \text{formamide}) - 600/n - 1.5(\% \text{mismatch}).$$

where C_i is the salt concentration (monovalent ions) and n is the length of the hybrid in base pairs (slightly modified from Meinkoth & Wahl (1984) *Anal. Biochem.* 138: 267-284).

In designing a hybridization experiment, some factors affecting nucleic acid hybridization can be conveniently altered. The temperature of the hybridization and washes and the salt concentration during the washes are the simplest to adjust. As the temperature of the hybridization increases (*ie.* stringency), it becomes less likely for hybridization to occur between strands that are nonhomologous, and as a result, background decreases. If the radiolabeled probe is not completely homologous with the immobilized fragment (as is frequently the case in gene family and interspecies hybridization experiments), the hybridization temperature must be reduced, and background will increase. The temperature of the washes affects the intensity of the hybridizing band and the degree of background in a similar manner. The stringency of the washes is also increased with decreasing salt concentrations.

In general, convenient hybridization temperatures in the presence of 50% formamide are 42°C for a probe with is 95% to 100% homologous to the target fragment, 37°C for 90% to 95% homology, and 32°C for 85% to 90% homology. For lower homologies, formamide content should be lowered and temperature adjusted accordingly, using the equation above. If the homology between the probe and the target fragment are not known, the simplest approach is to start with both hybridization and wash conditions which are nonstringent. If non-specific bands or high background are observed after autoradiography, the filter can be washed at high stringency and reexposed. If the time required for exposure makes this approach impractical, several hybridization and/or washing stringencies should be tested in parallel.

Nucleic Acid Probe Assays

Methods such as PCR, branched DNA probe assays, or blotting techniques utilizing nucleic acid probes according to the invention can determine the presence of cDNA or mRNA. A probe is said to “hybridize” with a sequence of the invention if it can form a duplex or double stranded complex, which is stable enough to be detected.

The nucleic acid probes will hybridize to the Neisserial nucleotide sequences of the invention (including both sense and antisense strands). Though many different nucleotide sequences will encode the amino acid sequence, the native Neisserial sequence is preferred because it is the actual sequence present in cells. mRNA represents a coding sequence and so a probe should be complementary to the coding sequence; single-stranded cDNA is complementary to mRNA, and so a cDNA probe should be complementary to the non-coding sequence.

The probe sequence need not be identical to the Neisserial sequence (or its complement) — some variation in the sequence and length can lead to increased assay sensitivity if the nucleic acid probe can form a duplex with target nucleotides, which can be detected. Also, the nucleic acid probe can include additional nucleotides to stabilize the formed duplex. Additional Neisserial sequence may also be helpful as a label to detect the formed duplex. For example, a non-complementary nucleotide sequence may be attached to the 5' end of the probe, with the remainder of the probe sequence being complementary to a Neisserial sequence. Alternatively, non-complementary bases or longer sequences can be interspersed into the probe, provided that the probe sequence has sufficient complementarity with the a Neisserial sequence in order to hybridize therewith and thereby form a duplex which can be detected.

The exact length and sequence of the probe will depend on the hybridization conditions, such as temperature, salt condition and the like. For example, for diagnostic applications, depending on the complexity of the analyte sequence, the nucleic acid probe typically contains at least 10-20 nucleotides, preferably 15-25, and more preferably at least 30 nucleotides, although it may be shorter than this. Short primers generally require cooler temperatures to form sufficiently stable hybrid complexes with the template.

Probes may be produced by synthetic procedures, such as the triester method of Matteucci *et al.* [*J. Am. Chem. Soc.* (1981) 103:3185], or according to Urdea *et al.* [*Proc. Natl. Acad. Sci. USA* (1983) 80: 7461], or using commercially available automated oligonucleotide synthesizers.

The chemical nature of the probe can be selected according to preference. For certain applications, DNA or RNA are appropriate. For other applications, modifications may be incorporated *eg.* backbone modifications, such as phosphorothioates or methylphosphonates, can be used to increase *in vivo* half-life, alter RNA affinity, increase nuclease resistance *etc.* [*eg.* see Agrawal & Iyer (1995) *Curr Opin Biotechnol* 6:12-19; Agrawal (1996) *TIBTECH* 14:376-387]; analogues such as peptide nucleic acids may also be used [*eg.* see Corey (1997) *TIBTECH* 15:224-229; Buchardt *et al.* (1993) *TIBTECH* 11:384-386].

Alternatively, the polymerase chain reaction (PCR) is another well-known means for detecting small amounts of target nucleic acids. The assay is described in: Mullis *et al.* [*Meth. Enzymol.* (1987) 155: 335-350]; US patents 4,683,195 and 4,683,202. Two "primer" nucleotides hybridize with the target nucleic acids and are used to prime the reaction. The primers can comprise sequence that does not hybridize to the sequence of the amplification target (or its complement) to aid with duplex stability or, for example, to incorporate a convenient restriction site. Typically, such sequence will flank the desired Neisserial sequence.

A thermostable polymerase creates copies of target nucleic acids from the primers using the original target nucleic acids as a template. After a threshold amount of target nucleic acids are generated by the polymerase, they can be detected by more traditional methods, such as Southern blots. When using the Southern blot method, the labelled probe will hybridize to the Neisserial sequence (or its complement).

Also, mRNA or cDNA can be detected by traditional blotting techniques described in Sambrook *et al* [*supra*]. mRNA, or cDNA generated from mRNA using a polymerase enzyme, can be purified and separated using gel electrophoresis. The nucleic acids on the gel are then blotted onto a solid support, such as nitrocellulose. The solid support is exposed to a labelled probe and then washed to remove any unhybridized probe. Next, the duplexes containing the labeled probe are detected. Typically, the probe is labelled with a radioactive moiety.

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1-20 show biochemical data obtained in the Examples, and also sequence analysis, for ORFs 37, 5, 2, 15, 22, 28, 32, 4, 61, 76, 89, 97, 106, 138, 23, 25, 27, 79, 85 and 132. M1 and M2 are molecular weight markers. Arrows indicate the position of the main recombinant product or, in Western blots, the position of the main *N.meningitidis* immunoreactive band. TP indicates *N.meningitidis* total protein extract; OMV indicates *N.meningitidis* outer membrane vesicle preparation. In bactericidal assay results: a diamond (♦) shows preimmune data; a triangle (▲) shows GST control data; a circle (●) shows data with recombinant *N.meningitidis* protein. Computer analyses show a hydrophilicity plot (upper), an antigenic index plot (middle), and an AMPHI analysis (lower). The AMPHI program has been used to predict T-cell epitopes [Gao *et al.* (1989) *J. Immunol.* **143**:3007; Roberts *et al.* (1996) *AIDS Res Hum Retrovir* **12**:593; Quakyi *et al.* (1992) *Scand J Immunol* suppl.11:9] and is available in the Protean package of DNASTAR, Inc. (1228 South Park Street, Madison, Wisconsin 53715 USA).

Figure 21 shows an alignment comparison of amino acid sequences for ORF 4 for several strains of *Neisseria*. Dark shading indicates regions of homology, and gray shading indicates the conservation of amino acids with similar characteristics. The Figure demonstrates a high degree of conservation among the various strains, further confirming its utility as an antigen for both vaccines and diagnostics.

EXAMPLES

The examples describe nucleic acid sequences which have been identified in *N.meningitidis*, along with their putative translation products, and also those of *N.gonorrhoeae*. Not all of the nucleic acid sequences are complete *ie.* they encode less than the full-length wild-type protein.

The examples are generally in the following format:

- a nucleotide sequence which has been identified in *N.meningitidis* (strain B)
- the putative translation product of this sequence
- a computer analysis of the translation product based on database comparisons
- corresponding gene and protein sequences identified in *N.meningitidis* (strain A) and in *N.gonorrhoeae*

- a description of the characteristics of the proteins which indicates that they might be suitably antigenic
- results of biochemical analysis (expression, purification, ELISA, FACS *etc.*)

The examples typically include details of sequence identity between species and strains. Proteins that are similar in sequence are generally similar in both structure and function, and the sequence identity often indicates a common evolutionary origin. Comparison with sequences of proteins of known function is widely used as a guide for the assignment of putative protein function to a new sequence and has proved particularly useful in whole-genome analyses.

Sequence comparisons were performed at NCBI (<http://www.ncbi.nlm.nih.gov>) using the algorithms BLAST, BLAST2, BLASTn, BLASTp, tBLASTn, BLASTx, & tBLASTx [*eg.* see also Altschul *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 25:2289-3402]. Searches were performed against the following databases: non-redundant GenBank+EMBL+DDBJ+PDB sequences and non-redundant GenBank CDS translations+PDB+SwissProt+SPupdate+PIR sequences.

To compare Meningococcal and Gonococcal sequences, the tBLASTx algorithm was used, as implemented at http://www.genome.ou.edu/gono_blast.html. The FASTA algorithm was also used to compare the ORFs (from GCG Wisconsin Package, version 9.0).

Dots within nucleotide sequences (*eg.* position 495 in SEQ ID 11) represent nucleotides which have been arbitrarily introduced in order to maintain a reading frame. In the same way, double-underlined nucleotides were removed. Lower case letters (*eg.* position 496 in SEQ ID 11) represent ambiguities which arose during alignment of independent sequencing reactions (some of the nucleotide sequences in the examples are derived from combining the results of two or more experiments).

Nucleotide sequences were scanned in all six reading frames to predict the presence of hydrophobic domains using an algorithm based on the statistical studies of Esposti *et al.* [Critical evaluation of the hydropathy of membrane proteins (1990) *Eur J Biochem* 190:207-219]. These domains represent potential transmembrane regions or hydrophobic leader sequences.

Open reading frames were predicted from fragmented nucleotide sequences using the program ORFFINDER (NCBI).

Underlined amino acid sequences indicate possible transmembrane domains or leader sequences in the ORFs, as predicted by the PSORT algorithm (<http://www.psort.nibb.ac.jp>). Functional domains were also predicted using the MOTIFS program (GCG Wisconsin & PROSITE).

Various tests can be used to assess the *in vivo* immunogenicity of the proteins identified in the examples. For example, the proteins can be expressed recombinantly and used to screen patient sera by immunoblot. A positive reaction between the protein and patient serum indicates that the patient has previously mounted an immune response to the protein in question *ie.* the protein is an immunogen. This method can also be used to identify immunodominant proteins.

The recombinant protein can also be conveniently used to prepare antibodies *eg.* in a mouse. These can be used for direct confirmation that a protein is located on the cell-surface. Labelled antibody (*eg.* fluorescent labelling for FACS) can be incubated with intact bacteria and the presence of label on the bacterial surface confirms the location of the protein.

In particular, the following methods (A) to (S) were used to express, purify and biochemically characterise the proteins of the invention:

A) Chromosomal DNA preparation

N.meningitidis strain 2996 was grown to exponential phase in 100ml of GC medium, harvested by centrifugation, and resuspended in 5ml buffer (20% Sucrose, 50mM Tris-HCl, 50mM EDTA, pH8).

After 10 minutes incubation on ice, the bacteria were lysed by adding 10ml lysis solution (50mM NaCl, 1% Na-Sarkosyl, 50µg/ml Proteinase K), and the suspension was incubated at 37°C for 2 hours. Two phenol extractions (equilibrated to pH 8) and one CHCl_3 /isoamylalcohol (24:1) extraction were performed. DNA was precipitated by addition of 0.3M sodium acetate and 2 volumes ethanol, and was collected by centrifugation. The pellet was washed once with 70% ethanol and redissolved in 4ml buffer (10mM Tris-HCl, 1mM EDTA, pH 8). The DNA concentration was measured by reading the OD at 260 nm.

B) Oligonucleotide design

Synthetic oligonucleotide primers were designed on the basis of the coding sequence of each ORF, using (a) the meningococcus B sequence when available, or (b) the gonococcus/meningococcus A sequence, adapted to the codon preference usage of meningococcus as necessary. Any predicted
 5 signal peptides were omitted, by deducing the 5'-end amplification primer sequence immediately downstream from the predicted leader sequence.

For most ORFs, the 5' primers included two restriction enzyme recognition sites (*Bam*HI-*Nde*I, *Bam*HI-*Nhe*I, or *Eco*RI-*Nhe*I, depending on the gene's own restriction pattern); the 3' primers included a *Xho*I restriction site. This procedure was established in order to direct the cloning of
 10 each amplification product (corresponding to each ORF) into two different expression systems: pGEX-KG (using either *Bam*HI-*Xho*I or *Eco*RI-*Xho*I), and pET21b+ (using either *Nde*I-*Xho*I or *Nhe*I-*Xho*I).

5'-end primer tail: CGCGGATCCCATATG (*Bam*HI-*Nde*I)

CGCGGATCCGCTAGC (*Bam*HI-*Nhe*I)

15 CCGGAATTCTAGCTAGC (*Eco*RI-*Nhe*I)

3'-end primer tail: CCCGCTCGAG (*Xho*I)

For ORFs 5, 15, 17, 19, 20, 22, 27, 28, 65 & 89, two different amplifications were performed to clone each ORF in the two expression systems. Two different 5' primers were used for each ORF; the same 3' *Xho*I primer was used as before:

20 5'-end primer tail: GGAATTCCATATGGCCATGG (*Nde*I)

5'-end primer tail: CGGGATCC (*Bam*HI)

ORF 76 was cloned in the pTRC expression vector and expressed as an amino-terminus His-tag fusion. In this particular case, the predicted signal peptide was included in the final product. *Nhe*I-*Bam*HI restriction sites were incorporated using primers:

25 5'-end primer tail: GATCAGCTAGCCATATG (*Nhe*I)

3'-end primer tail: CGGGATCC (*Bam*HI)

As well as containing the restriction enzyme recognition sequences, the primers included nucleotides which hybridized to the sequence to be amplified. The number of hybridizing nucleotides depended on the melting temperature of the whole primer, and was determined for each primer using the formulae:

$$T_m = 4 (G+C) + 2 (A+T) \quad (\text{tail excluded})$$

$$T_m = 64.9 + 0.41 (\% \text{ GC}) - 600/N \quad (\text{whole primer})$$

The average melting temperature of the selected oligos were 65-70°C for the whole oligo and 50-55°C for the hybridising region alone.

Table I shows the forward and reverse primers used for each amplification. In certain cases, it will be noted that the sequence of the primer does not exactly match the sequence in the ORF. When initial amplifications were performed, the complete 5' and/or 3' sequence was not known for some meningococcal ORFs, although the corresponding sequences had been identified in gonococcus. For amplification, the gonococcal sequences could thus be used as the basis for primer design, altered to take account of codon preference. In particular, the following codons were changed: ATA→ATT; TCG→TCT; CAG→CAA; AAG→AAA; GAG→GAA; CGA→CGC; CGG→CGC; GGG→GGC. Italicised nucleotides in Table I indicate such a change. It will be appreciated that, once the complete sequence has been identified, this approach is generally no longer necessary.

TABLE I – PCR primers

ORF	Primer	Sequence	Restriction sites
ORF 1	Forward	CGCGGATCCGCTAGC-GGACACACTTATTTCGG <SEQ ID 924>	BamHI-NheI
	Reverse	CCCGCTCGAG-CCAGCGGTAGCCTAATT <SEQ ID 925>	XhoI
ORF 2	Forward	GCGGATCCCATATG-TTTGATTTCGGTTTGGG <SEQ ID 926>	BamHI-NdeI
	Reverse	CCCGCTCGAG-GACGGCATAACGGCG <SEQ ID 927>	XhoI
ORF 2-1	Forward	GCGGATCCCATATG-TTTGATTTCGGTTTGGG <SEQ ID 928>	BamHI-NdeI
	Reverse	CCCGCTCGAG-TGATTACGGACGCGCA <SEQ ID 929>	XhoI
ORF 4	Forward	GCGGATCCCATATG-TGCGGAGGTCAAAAAGAC <SEQ ID 930>	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTTGGCTGCGCCTTC <SEQ ID 931>	XhoI

ORF 5	Forward	GGAATTCCATATG <u>GC</u> CCATGG-TGGAAGCGGCACAACC <SEQ ID 932>	NdeI-NcoI
	Forward	CGGGATCC-ATGGAAGGCGCACAAAC <SEQ ID 933>	BamHI
	Reverse	CCCGCTCGAG-GACTGTGCAAAACCG <SEQ ID 934>	XhoI
ORF 6	Forward	CGCGGATCCCATATG-ACCCGCTCAATCTCTGCA <SEQ ID 935>	BamHI-NdeI
	Reverse	CCCGCTCGAG-TGCGCGGAACACTTTC <SEQ ID 936>	XhoI
ORF 7	Forward	CGCGGATCCGCTAGC-GCGCTGCTTTTGTTC <SEQ ID 937>	BamHI-NheI
	Reverse	CCCGCTCGAG-TTTCAAAATATATTGCGGA <SEQ ID 938>	XhoI
ORF 8	Forward	GCGGATCCCATATG-GCTCAACTGCTTCGTAC <SEQ ID 939>	BamHI-NdeI
	Reverse	CCCGCTCGAG-AGCAGGCTTTGCGCG <SEQ ID 940>	XhoI
ORF 9	Forward	CGCGGATCCCATATG-CCGAAGGAAGTCGAAA <SEQ ID 941>	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTTCGAGGTTTTCGGG <SEQ ID 942>	XhoI
ORF 10	Forward	GCGGATCCCATATG-GACACAAAAGAAATCCTC <SEQ ID 943>	BamHI-NdeI
	Reverse	CCCGCTCGAG-TAATGGGAACCTTGTTTT <SEQ ID 944>	XhoI
ORF 11	Forward	GCGGATCCCATATG-GCGGTCAACCTCTACG <SEQ ID 945>	BamHI-NdeI
	Reverse	CCCGCTCGAG-GGAAACGACTTCGCC <SEQ ID 946>	XhoI
ORF 13	Forward	CGCGGATCCCATATG-GCTCTGCTTTCGCGC <SEQ ID 947>	BamHI-NdeI
	Reverse	CCCGCTCGAG-AGGGTGTGTGATAATAAG <SEQ ID 948>	XhoI
ORF 15	Forward	GGAATTCATATGGCCATGG-GCGGGACACTGACAG <SEQ ID 949>	NdeI-NcoI
	Forward	CGGGATCC-TGCGGGGACACTGACAGG <SEQ ID 950>	BamHI
	Reverse	CCCGCTCGAG-AGGTTGGCCTTGTCATG <SEQ ID 951>	XhoI
ORF 17	Forward	GGAATTCATATGGCCATGG-TTGCCGGCCTGTTCG <SEQ ID 952>	NdeI-NcoI
	Forward	CGGGATCC-ATTGCCGGCCTGTTCG <SEQ ID 953>	BamHI
	Reverse	CCCGCTCGAG-AAGCAGGTTGTACAGC <SEQ ID 954>	XhoI
ORF 18	Forward	GCGGATCCCATATG-ATTTTGCTGCATTTGGAT <SEQ ID 955>	BamHI-NdeI
	Reverse	CCCGCTCGAG-TCTTCCAATTCTGAAAGC <SEQ ID 956>	XhoI
ORF 19	Forward	GGAATTCATATGGCCATGG-TCGCCAGTGTTTTACC <SEQ ID 957>	NdeI-NcoI
	Forward	CGGGATCC-TTCGCCAGTGTTTTTACCG <SEQ ID 958>	BamHI
	Reverse	CCCGCTCGAG-GGTGTTTTTGAAGCTGCC <SEQ ID 959>	XhoI
ORF 20	Forward	GGAATTCATATGGCCATGG-TCGGCGCGGGTATG <SEQ ID 960>	NdeI-NcoI

ORF 22	Forward	CGGGATCC-TTCGGCGCGGGTATG <SEQ ID 961>	BamHI
	Reverse	CCCGCTCGAG-CGGCGAGCGAGAGCA <SEQ ID 962>	XhoI
ORF 22	Forward	GGAATTCATATGGCCATGG-TGATTAATAATCAAAAAAGGTCT <SEQ ID 963>	NdeI-NcoI
	Forward	CGGGATCC-ATGATTAATAATCAAAAAAGGTCTAAACC <SEQ ID 964>	BamHI
	Reverse	CCCGCTCGAG-ATTATGATAGCGGCC <SEQ ID 965>	XhoI
ORF 23	Forward	CGCGGATCCCATATG-GATGTTTCTGTTTCAGAC <SEQ ID 966>	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTTAAACCGATAGGTAAACG <SEQ ID 967>	XhoI
ORF 24	Forward	GGAATTCATATGGCCATGG -TGATGCCGGAATGGTG <SEQ ID 968>	NdeI-NcoI
	Forward	CGGGATCC-ATGATGCCGGAATGGTG <SEQ ID 969>	BamHI
	Reverse	CCCGCTCGAG-TGTCAGCGTGGCGCA <SEQ ID 970>	XhoI
ORF 25	Forward	GCGGATCCCATATG-TATCGCAAATGATTC <SEQ ID 971>	BamHI-NdeI
	Reverse	CCCGCTCGAG-ATCGATGGAATAGCCG <SEQ ID 972>	XhoI
ORF 26	Forward	GCGGATCCCATATG -CAGCTGATCGACTATTC <SEQ ID 973>	BamHI-NdeI
	Reverse	CCCGCTCGAG-GACATCGGCGCGTTTT <SEQ ID 974>	XhoI
ORF 27	Forward	GGAATTCATATGGCCATGG-AGACCTATTCTGTTTA <SEQ ID 974>	NdeI-NcoI
	Forward	CGGGATCC- CAGACCTATTCTGTTTATTTTAATC <SEQ ID 975>	BamHI
	Reverse	CCCGCTCGAG-GGGTTCGATTAAATAACCAT <SEQ ID 976>	XhoI
ORF 28	Forward	GGAATTCATATGGCCATGG-ACGGCTGTACGTTGATGT <SEQ ID 977>	NdeI-NcoI
	Forward	CGGGATCC-AACGGCTGTACGTTGATG <SEQ ID 978>	BamHI
	Reverse	CCCGCTCGAG-TTTGTCAGAGGAATTCGCG <SEQ ID 979>	XhoI
ORF 29	Forward	GCGGATCCCATATG -AACGGTTTGATGCCCG <SEQ ID 980>	BamHI-NdeI
	Forward	CGCGGATCCGCTAGC-AACGGTTTGATGCCCG <SEQ ID 981>	BamHI-NheI
	Reverse	CCCGCTCGAG-TTGTGCTAAGTTCCTGATATG <SEQ ID 982>	XhoI
ORF 32	Forward	CGCGGATCCCATATG-AATACTCCTCCTTTTG <SEQ ID 983>	BamHI-NdeI
	Reverse	CCCGCTCGAG-GCGTATTTTTTGATGCTTTG <SEQ ID 984>	XhoI
ORF 33	Forward	GCGGATCCCATATG -ATTGATAGGGATCGTATG <SEQ ID 985>	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTGATCTTTCAAACGGCC <SEQ ID 986>	XhoI
ORF 35	Forward	GCGGATCCCATATG-TTCAGAGCTCAGCTT <SEQ ID 987>	BamHI-NdeI
	Forward	CCCGGATCCGCTAGC-TTCAGAGCTCAGCTT <SEQ ID 988>	BamHI-NheI
	Reverse	CCCGCTCGAG-AAACAGCCATTGAGCGA <SEQ ID 989>	XhoI

ORF 37	Forward	CGGGATCCCATATG-GATGACGTATCGGATTTT <SEQ ID 990>	BamHI-NdeI
	Reverse	CCCGCTCGAG-ATAGCCCGCTTTTCAGG <SEQ ID 991>	XhoI
ORF 58	Forward	CGCGGATCCGCTAGC-TCCGAACGCGAGTGGAT <SEQ ID 992>	BamHI-NheI
	Reverse	CCCGCTCGAG-AGCATTTGTCCAAGGGGAC <SEQ ID 993>	XhoI
ORF 65	Forward	GGAATTCATATGGCCATGG -TGCTGTATCTGAATCAAG <SEQ ID 994>	NdeI-NcoI
	Forward	CGGGATCC-TTGCTGTATCTGAATCAAG <SEQ ID 995>	BamHI
	Reverse	CCCGCTCGAG-CCGCATCGGCAGACA <SEQ ID 996>	XhoI
ORF 66	Forward	GCGGATCCCATATG-TACGCATTTACGCCG <SEQ ID 997>	BamHI-NdeI
	Reverse	CCCGCTCGAG-TGGATTTTGACAGATGG <SEQ ID 998>	XhoI
ORF 72	Forward	CGCGGATCCCATATG- AATGCAAGAAAAATATCTGA <SEQ ID 999>	BamHI-NdeI
	Reverse	CCCGCTCGAG-GCCTGAGACCTTTGCAA <SEQ ID 1000>	XhoI
ORF 73	Forward	GCGGATCCCATATG-AGATTTTTCGGTATCGG <SEQ ID 1001>	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTCATCTTTTCATGTTTCG <SEQ ID 1002>	XhoI
ORF 75	Forward	GCGGATCCCATATG- TCTGTCTTTCAAACGGC <SEQ ID 1003>	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTTGTTTTTGCAAGACAG <SEQ ID 1004>	XhoI
ORF 76	Forward	GATCAGCTAGCCATATG-AAACAGAAAAAACCGC <SEQ ID 1005>	NheI-NdeI
	Reverse	CGGGATCC-TTACGTTTGACACCGTT <SEQ ID 1006>	BamHI
ORF 79	Forward	CGCGGATCCCATATG-GTTTCCGCCGCCG <SEQ ID 1007>	BamHI-NdeI
	Reverse	CCCGCTCGAG-GTGCTGATGCGCTTCG <SEQ ID 1008>	XhoI
ORF 83	Forward	GCGGATCCCATATG-AAAACCTGTGCTGTC <SEQ ID 1009>	BamHI-NdeI
	Reverse	CCCGCTCGAG-GCCGCCTTTGCGGC <SEQ ID 1010>	XhoI
ORF 84	Forward	GCGGATCCCATATG-GCAGAGATCTGTTTG <SEQ ID 1011>	BamHI-NdeI
	Reverse	CCCGCTCGAG-GTTTGCCGATCCGACCA <SEQ ID 1012>	XhoI
ORF 85	Forward	CGCGGATCCCATATG- CGCGTTTGGGGCGGA <SEQ ID 1013>	BamHI-NdeI
	Reverse	CCCGCTCGAG-TCGGCGCGCGGGC <SEQ ID 1014>	XhoI
ORF 89	Forward	GGAATTCATATGGCCATGG-CCATACCTTCTTATCA <SEQ ID 1015>	NdeI-NcoI
	Forward	CGGGATCC-GCCATACCTTCTTATCAGAG <SEQ ID 1016>	BamHI
	Reverse	CCCGCTCGAG-TTTTTGCGATTAGAAAAAGC <SEQ ID 1017>	XhoI

ORF 97	Forward Reverse	GCGGATCCCATATG-CATCCTGCCAGCGAAC <SEQ ID 1018> CCCGCTCGAG-TTCGCCTACGTTTTTTG <SEQ ID 1019>	BamHI-NdeI XhoI
ORF 98	Forward Reverse	GCGGATCCCATATG-ACGGTAACTGCGG <SEQ ID 1020> CCCGCTCGAG-TTGTGTTCGGGCAATC <SEQ ID 1021>	BamHI-NdeI XhoI
ORF 100	Forward Reverse	GCGGATCCCATATG-TCGGGCATTTACACCG <SEQ ID 1022> CCCGCTCGAG-ACGGGTTTCGGCGGAA <SEQ ID 1023>	BamHI-NdeI XhoI
ORF 101	Forward Reverse	GCGGATCCCATATG-ATTTATCAAGAAACCTC <SEQ ID 1024> CCCGCTCGAG-TTTCCGCCTTTCAATGT <SEQ ID 1025>	BamHI-NdeI XhoI
ORF 102	Forward Reverse	GCGGATCCCATATG-GCAGGGCTGTTTTACC <SEQ ID 1026> CCCGCTCGAG-AAACGGTTTGAACACGAC <SEQ ID 1027>	BamHI-NdeI XhoI
ORF 103	Forward Reverse	GCGGATCCCATATG-AACCACGACATCAC <SEQ ID 1028> CCCGCTCGAG-CAGCCAAGGACGGC <SEQ ID 1029>	BamHI-NdeI XhoI
ORF 104	Forward Reverse	GCGGATCCCATATG-ACGTGGGGAACGC <SEQ ID 1030> CCCGCTCGAG-GCGGCGTTGAACGGC <SEQ ID 1031>	BamHI-NdeI XhoI
ORF 105	Forward Reverse	GCGGATCCCATATG-ACCAAATTTCAAACCCCTC <SEQ ID 1032> CCCGCTCGAG-TAAACGAATGCCGTCCAG <SEQ ID 1033>	BamHI-NdeI XhoI
ORF 106	Forward Reverse	GCGGATCCCATATG-AGGATAAACCAGCGCG <SEQ ID 1034> CCCGCTCGAG-TTGTGTCCCGATGATGT <SEQ ID 1035>	BamHI-NdeI XhoI
ORF 109	Forward Reverse	GCGGATCCCATATG-GAAGATTTATATATAACTCTG <SEQ ID 1036> CCCGCTCGAG-ATCAGCTTCGAACCGAAG <SEQ ID 1037>	BamHI-NdeI XhoI
ORF110	Forward Reverse	AAAGAATTC-ATGAGTAAATCCGTAGATCTCCC <SEQ ID 1038> AAACTGCAG-GGAAAACCACATCCGCACTCTGCC <SEQ ID 1039>	EcoRI PstI
ORF111	Forward Reverse	AAAGAATTC-GCACCGCAAAAGGCAAAACCGCA <SEQ ID 1040> AAACTGCAG-TCTGCGCGTTTTCGGGCAGGGTGG <SEQ ID 1041>	EcoRI PstI
ORF113	Forward Reverse	AAAGAATTC-ATGAACAAAACCTCTATCGTGTGATTTCAACCG <SEQ ID 1042> AAACTGCAG-TTACGAATGCTGCTTCTCGACCGTACTG <SEQ ID 1043>	EcoRI PstI
ORF115	Forward Reverse	AAAGAATTC-TTGCTTGTGCAAAACAGAAAAGACGG <SEQ ID 1044> AAAAAGTCGAC-CTATTTTTTAGGGGCTTTGCTTGTGTTGAAAAGCTGCC <SEQ ID 1045>	EcoRI SalI

ORF119	Forward	<u>AAAGAATTC</u> -TACAACATGTATCAGGAAAACCAATACCG <SEQ ID 1046>	EcoRI
	Reverse	AAACTGCAG-TTATGAAAACAGGCGCAGGCGGTTTGCC <SEQ ID 1047>	PstI
ORF120	Forward	<u>AAAGAATTC</u> -GCAAGGCTACCCCAATCCGCCGTG <SEQ ID 1048>	EcoRI
	Reverse	<u>AAACTGCAG</u> -CGGTTTGGCTGCCTGGCCGTGAT <SEQ ID 1049>	PstI
ORF121	Forward	<u>AAAGAATTC</u> -GCCTTGGTCTGGCTGGTTTTTCGC <SEQ ID 1050>	EcoRI
	Reverse	<u>AAACTGCAG</u> -TCATCCGCCACCCACCTCGGCCATCCATC <SEQ ID 1051>	PstI
ORF122	Forward	<u>AAAAAGTCGAC</u> -ATGCTTTACCGCGCAAGCAGTTC <SEQ ID 1052>	SalI
	Reverse	<u>AAACTGCAG</u> -TCAGGAACACAAACGATGACGAATATCCGTATC <SEQ ID 1053>	PstI
ORF125	Forward	<u>AAAGAATTC</u> -GCGCTGTTTTTTCGCGCGGCGTAT <SEQ ID 1054>	EcoRI
	Reverse	<u>AAACTGCAG</u> -CGCCGTTTCAAGACGAAAAAGTCG <SEQ ID 1055>	PstI
ORF126	Forward	<u>AAAGAATTC</u> -GCGGAACCGGTCGAAG <SEQ ID 1056>	EcoRI
	Reverse	<u>AAACTGCAG</u> -TTAATCTTGCTTCCGATATAC <SEQ ID 1057>	PstI
ORF127	Forward	<u>AAAGAATTC</u> -ATGACTGATAATCGGGGTTTACG <SEQ ID 1058>	EcoRI
	Reverse	<u>AAAAAGTCGAC</u> -CTTAAGTAACTGCAGTCCTTATC <SEQ ID 1059>	SalI
ORF128	Forward	<u>AAAGAATTC</u> -ATGCAAGCTGTCCGCTACAGGCC <SEQ ID 1060>	EcoRI
	Reverse	<u>AAACTGCAG</u> -CTATTGCAATGCGCGCCCGGGGAATGTTGAGCAGGCG <SEQ ID 1061>	PstI
ORF129	Forward	<u>AAAGAATTC</u> -ATGGATTTTCGTTTTCGACATTATTACGAATACCG <SEQ ID 1062>	EcoRI
	Reverse	<u>AAACTGCAG</u> -TTATTTTTTGATGAAATTTTGGGGCGG <SEQ ID 1063>	PstI
ORF130	Forward	<u>AAAGAATTC</u> -GCAGTACTTGCCATTCTCGGTGCG <SEQ ID 1064>	EcoRI
	Reverse	<u>AAACTGCAG</u> -CTCCGGATCGTCTGTAAACGCATT <SEQ ID 1065>	PstI
ORF 131	Forward	<u>GCGGATCCCATATG</u> -GAAATTCGGGCAATAAAAT <SEQ ID 1066>	BamHI-NdeI
	Reverse	<u>CCCGCTCGAG</u> -CCAGCGGACGCGTTC <SEQ ID 1067>	XhoI
ORF 132	Forward	<u>GCGGATCCCATATG</u> -AAAGAAGCGGGGTTTG <SEQ ID 1068>	BamHI-NdeI
	Reverse	<u>CCCGCTCGAG</u> -CCAATCTGCCAGCGCT <SEQ ID 1069>	XhoI
ORF 133	Forward	<u>GCGGATCCCATATG</u> -GAAGATGCAGGCGCGC <SEQ ID 1070>	BamHI-NdeI

	Reverse	CCCGCTCGAG-AACTTGTAGCTCATCGT <SEQ ID 1071>	XhoI
ORF 134	Forward	GCGGATCCCATATG-TCTGTGCAAGCAGTATTG <SEQ ID 1072>	BamHI-NdeI
	Reverse	CCCGCTCGAG-ATCCTGTGCCAATGCG <SEQ ID 1073>	XhoI
ORF 135	Forward	GCGGATCCCATATG-CCGTCTGAAAAAGCTTT <SEQ ID 1074>	BamHI-NdeI
	Reverse	CCCGCTCGAG-AAATACCGCTGAGGATG <SEQ ID 1075>	XhoI
ORF 136	Forward	GCGGATCCGCTAGC-ATGAAGCGGCGTATAGCC <SEQ ID 1076>	BamHI-NheI
	Reverse	CCCGCTCGAG-TTCCGAATATTGGAACTTTT <SEQ ID 1077>	XhoI
ORF 137	Forward	GCGGATCCCATATG-GGCACGGCGGAAATA <SEQ ID 1078>	BamHI-NdeI
	Reverse	CCCGCTCGAG-ATAACGGTATGCCGCC <SEQ ID 1079>	XhoI
ORF 138	Forward	GCGGATCCCATATG-TTTCGTTTACAATTCAGGC <SEQ ID 1080>	BamHI-NdeI
	Reverse	CCCGCTCGAG-CGGCGTTTTATAGCG <SEQ ID 1081>	XhoI
ORF 139	Forward	GCGGATCCCATATG-GCTTTTGTGGCGTAATG <SEQ ID 1082>	BamHI-NdeI
	Reverse	CCCGCTCGAG-TAACGTTCCGTGCGTTT <SEQ ID 1083>	XhoI
ORF 140	Forward	GCGGATCCCATATG-TTGCCACAGGCAGC <SEQ ID 1084>	BamHI-NdeI
	Reverse	CCCGCTCGAG-GACGATGGCAACAGC <SEQ ID 1085>	XhoI
ORF 141	Forward	GCGGATCCCATATG-CCGTCTGAAGCAGTCT <SEQ ID 1086>	BamHI-NdeI
	Reverse	CCCGCTCGAG-ATCTGTTGTTTTTAAATATT <SEQ ID 1087>	XhoI
ORF 142	Forward	GCGGATCCCATATG-GATAATCTGTAGTGAAG <SEQ ID 1088>	BamHI-NdeI
	Reverse	CCCGCTCGAG-AAACGTATAGCCTACCT <SEQ ID 1089>	XhoI
ORF 143	Forward	GCGGATCCCATATG-GATACCGCTTTGAACCT <SEQ ID 1090>	BamHI-NdeI
	Reverse	CCCGCTCGAG-AATGGCTTCGCAATATG <SEQ ID 1091>	XhoI
ORF 144	Forward	GCGGATCCCATATG-ACCTTTTTACAACGTTTGC <SEQ ID 1092>	BamHI-NdeI
	Reverse	CCCGCTCGAG-AGATTGTTGTTGTTTTTCG <SEQ ID 1093>	XhoI
ORF 147	Forward	GCGGATCCCATATG-TCTGTCTTCAACAGCGC <SEQ ID 1094>	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTTGTTTTGCAAGACAG <SEQ ID 1095>	XhoI

NB:

- restriction sites are underlined

- for ORFs 110-130, where the ORF itself carries an *EcoRI* site (eg. ORF122), a *SaI* site was used in the forward primer instead. Similarly, where the ORF carries a *PstI* site (eg. ORFs 115 and 127), a *SaI* site was used in the reverse primer.

- 5 Oligos were synthesized by a Perkin Elmer 394 DNA/RNA Synthesizer, eluted from the columns in 2ml NH_4OH , and deprotected by 5 hours incubation at 56°C . The oligos were precipitated by addition of 0.3M Na-Acetate and 2 volumes ethanol. The samples were then centrifuged and the pellets resuspended in either 100 μl or 1ml of water. OD_{260} was determined using a Perkin Elmer Lambda Bio spectrophotometer and the concentration was determined and adjusted to 2-10pmol/ μl .

C) Amplification

- 10 The standard PCR protocol was as follows: 50-200ng of genomic DNA were used as a template in the presence of 20-40 μM of each oligo, 400-800 μM dNTPs solution, 1x PCR buffer (including 1.5mM MgCl_2), 2.5 units *TaqI* DNA polymerase (using Perkin-Elmer AmpliTaq, GIBCO Platinum, Pwo DNA polymerase, or Tahara Shuzo Taq polymerase).

In some cases, PCR was optimised by the addition of 10 μl DMSO or 50 μl 2M betaine.

- 15 After a hot start (adding the polymerase during a preliminary 3 minute incubation of the whole mix at 95°C), each sample underwent a double-step amplification: the first 5 cycles were performed using as the hybridization temperature the one of the oligos excluding the restriction enzymes tail, followed by 30 cycles performed according to the hybridization temperature of the whole length oligos. The cycles were followed by a final 10 minute extension step at 72°C .
- 20 The standard cycles were as follows:

	Denaturation	Hybridisation	Elongation
First 5 cycles	30 seconds 95°C	30 seconds $50-55^\circ\text{C}$	30-60 seconds 72°C
Last 30 cycles	30 seconds 95°C	30 seconds $65-70^\circ\text{C}$	30-60 seconds 72°C

The elongation time varied according to the length of the ORF to be amplified.

The amplifications were performed using either a 9600 or a 2400 Perkin Elmer GeneAmp PCR System. To check the results, 1/10 of the amplification volume was loaded onto a 1-1.5% agarose gel and the size of each amplified fragment compared with a DNA molecular weight marker.

The amplified DNA was either loaded directly on a 1% agarose gel or first precipitated with ethanol and resuspended in a suitable volume to be loaded on a 1% agarose gel. The DNA fragment corresponding to the right size band was then eluted and purified from gel, using the Qiagen Gel Extraction Kit, following the instructions of the manufacturer. The final volume of the DNA fragment was 30µl or 50µl of either water or 10mM Tris, pH 8.5.

D) Digestion of PCR fragments

The purified DNA corresponding to the amplified fragment was split into 2 aliquots and double-digested with:

- *NdeI/XhoI* or *NheI/XhoI* for cloning into pET-21b+ and further expression of the protein as a C-terminus His-tag fusion
- *BamHI/XhoI* or *EcoRI/XhoI* for cloning into pGEX-KG and further expression of the protein as N-terminus GST fusion.
- For ORF 76, *NheI/BamHI* for cloning into pTRC-HisA vector and further expression of the protein as N-terminus His-tag fusion.
- *EcoRI/PstI*, *EcoRI/SalI*, *Sall/PstI* for cloning into pGex-His and further expression of the protein as N-terminus His-tag fusion

Each purified DNA fragment was incubated (37°C for 3 hours to overnight) with 20 units of each restriction enzyme (New England Biolabs) in either 30 or 40µl final volume in the presence of the appropriate buffer. The digestion product was then purified using the QIAquick PCR purification kit, following the manufacturer's instructions, and eluted in a final volume of 30 or 50µl of either water or 10mM Tris-HCl, pH 8.5. The final DNA concentration was determined by 1% agarose gel electrophoresis in the presence of titrated molecular weight marker.

E) Digestion of the cloning vectors (pET22B, pGEX-KG, pTRC-His A, and pGex-His)

10µg plasmid was double-digested with 50 units of each restriction enzyme in 200µl reaction volume in the presence of appropriate buffer by overnight incubation at 37°C. After loading the whole digestion on a 1% agarose gel, the band corresponding to the digested vector was purified from the gel using the Qiagen QIAquick Gel Extraction Kit and the DNA was eluted in 50µl of 10mM Tris-HCl, pH 8.5. The DNA concentration was evaluated by measuring OD₂₆₀ of the sample, and adjusted to 50µg/µl. 1µl of plasmid was used for each cloning procedure.

The vector pGEX-His is a modified pGEX-2T vector carrying a region encoding six histidine residues upstream to the thrombin cleavage site and containing the multiple cloning site of the vector pTRC99 (Pharmacia).

F) Cloning

The fragments corresponding to each ORF, previously digested and purified, were ligated in both pET22b and pGEX-KG. In a final volume of 20µl, a molar ratio of 3:1 fragment/vector was ligated using 0.5µl of NEB T4 DNA ligase (400 units/µl), in the presence of the buffer supplied by the manufacturer. The reaction was incubated at room temperature for 3 hours. In some experiments, ligation was performed using the Boheringer "Rapid Ligation Kit", following the manufacturer's instructions.

In order to introduce the recombinant plasmid in a suitable strain, 100µl *E. coli* DH5 competent cells were incubated with the ligase reaction solution for 40 minutes on ice, then at 37°C for 3 minutes, then, after adding 800µl LB broth, again at 37°C for 20 minutes. The cells were then centrifuged at maximum speed in an Eppendorf microfuge and resuspended in approximately 200µl of the supernatant. The suspension was then plated on LB ampicillin (100mg/ml).

The screening of the recombinant clones was performed by growing 5 randomly-chosen colonies overnight at 37°C in either 2ml (pGEX or pTC clones) or 5ml (pET clones) LB broth + 100µg/ml ampicillin. The cells were then pelleted and the DNA extracted using the Qiagen QIAprep Spin Miniprep Kit, following the manufacturer's instructions, to a final volume of 30µl. 5µl of each individual miniprep (approximately 1g) were digested with either *NdeI/XhoI* or *BamHI/XhoI* and the whole digestion loaded onto a 1-1.5% agarose gel (depending on the expected insert size), in

parallel with the molecular weight marker (1Kb DNA Ladder, GIBCO). The screening of the positive clones was made on the base of the correct insert size.

For the cloning of ORFs 110, 111, 113, 115, 119, 122, 125 & 130, the double-digested PCR product was ligated into double-digested vector using *EcoRI-PstI* cloning sites or, for ORFs 115 & 127, *EcoRI-SalI* or, for ORF 122, *Sall-PstI*. After cloning, the recombinant plasmids were introduced in the *E.coli* host W3110. Individual clones were grown overnight at 37°C in L-broth with 50µl/ml ampicillin.

G) Expression

Each ORF cloned into the expression vector was transformed into the strain suitable for expression of the recombinant protein product. 1µl of each construct was used to transform 30µl of *E.coli* BL21 (pGEX vector), *E.coli* TOP 10 (pTRC vector) or *E.coli* BL21-DE3 (pET vector), as described above. In the case of the pGEX-His vector, the same *E.coli* strain (W3110) was used for initial cloning and expression. Single recombinant colonies were inoculated into 2ml LB+Amp (100µg/ml), incubated at 37°C overnight, then diluted 1:30 in 20ml of LB+Amp (100µg/ml) in 100ml flasks, making sure that the OD₆₀₀ ranged between 0.1 and 0.15. The flasks were incubated at 30°C into gyratory water bath shakers until OD indicated exponential growth suitable for induction of expression (0.4-0.8 OD for pET and pTRC vectors; 0.8-1 OD for pGEX and pGEX-His vectors). For the pET, pTRC and pGEX-His vectors, the protein expression was induced by addition of 1mM IPTG, whereas in the case of pGEX system the final concentration of IPTG was 0.2mM. After 3 hours incubation at 30°C, the final concentration of the sample was checked by OD. In order to check expression, 1ml of each sample was removed, centrifuged in a microfuge, the pellet resuspended in PBS, and analysed by 12% SDS-PAGE with Coomassie Blue staining. The whole sample was centrifuged at 6000g and the pellet resuspended in PBS for further use.

H) GST-fusion proteins large-scale purification.

A single colony was grown overnight at 37°C on LB+Amp agar plate. The bacteria were inoculated into 20ml of LB+Amp liquid culture in a water bath shaker and grown overnight. Bacteria were diluted 1:30 into 600ml of fresh medium and allowed to grow at the optimal temperature (20-37°C) to OD₅₅₀ 0.8-1. Protein expression was induced with 0.2mM IPTG followed by three hours incubation. The culture was centrifuged at 8000rpm at 4°C. The supernatant was discarded and the

bacterial pellet was resuspended in 7.5ml cold PBS. The cells were disrupted by sonication on ice for 30 sec at 40W using a Branson sonifier B-15, frozen and thawed twice and centrifuged again. The supernatant was collected and mixed with 150 μ l Glutathione-Sepharose 4B resin (Pharmacia) (previously washed with PBS) and incubated at room temperature for 30 minutes. The sample was centrifuged at 700g for 5 minutes at 4°C. The resin was washed twice with 10ml cold PBS for 10 minutes, resuspended in 1ml cold PBS, and loaded on a disposable column. The resin was washed twice with 2ml cold PBS until the flow-through reached OD₂₈₀ of 0.02-0.06. The GST-fusion protein was eluted by addition of 700 μ l cold Glutathione elution buffer (10mM reduced glutathione, 50mM Tris-HCl) and fractions collected until the OD₂₈₀ was 0.1. 21 μ l of each fraction were loaded on a 12% SDS gel using either Biorad SDS-PAGE Molecular weight standard broad range (M1) (200, 116.25, 97.4, 66.2, 45, 31, 21.5, 14.4, 6.5 kDa) or Amersham Rainbow Marker (M2) (220, 66, 46, 30, 21.5, 14.3 kDa) as standards. As the MW of GST is 26kDa, this value must be added to the MW of each GST-fusion protein.

I) His-fusion solubility analysis (ORFs 111-129)

To analyse the solubility of the His-fusion expression products, pellets of 3ml cultures were resuspended in buffer M1 [500 μ l PBS pH 7.2]. 25 μ l lysozyme (10mg/ml) was added and the bacteria were incubated for 15 min at 4°C. The pellets were sonicated for 30 sec at 40W using a Branson sonifier B-15, frozen and thawed twice and then separated again into pellet and supernatant by a centrifugation step. The supernatant was collected and the pellet was resuspended in buffer M2 [8M urea, 0.5M NaCl, 20mM imidazole and 0.1M NaH₂PO₄] and incubated for 3 to 4 hours at 4°C. After centrifugation, the supernatant was collected and the pellet was resuspended in buffer M3 [6M guanidinium-HCl, 0.5M NaCl, 20mM imidazole and 0.1M NaH₂PO₄] overnight at 4°C. The supernatants from all steps were analysed by SDS-PAGE.

The proteins expressed from ORFs 113, 119 and 120 were found to be soluble in PBS, whereas ORFs 111, 122, 126 and 129 need urea and ORFs 125 and 127 need guanidium-HCl for their solubilization.

J) His-fusion large-scale purification.

A single colony was grown overnight at 37°C on a LB + Amp agar plate. The bacteria were inoculated into 20ml of LB+Amp liquid culture and incubated overnight in a water bath shaker.

Bacteria were diluted 1:30 into 600ml fresh medium and allowed to grow at the optimal temperature (20-37°C) to OD₅₅₀ 0.6-0.8. Protein expression was induced by addition of 1mM IPTG and the culture further incubated for three hours. The culture was centrifuged at 8000rpm at 4°C, the supernatant was discarded and the bacterial pellet was resuspended in 7.5ml of either (i) cold
5 buffer A (300mM NaCl, 50mM phosphate buffer, 10mM imidazole, pH 8) for soluble proteins or (ii) buffer B (urea 8M, 10mM Tris-HCl, 100mM phosphate buffer, pH 8.8) for insoluble proteins.

The cells were disrupted by sonication on ice for 30 sec at 40W using a Branson sonifier B-15, frozen and thawed two times and centrifuged again.

For insoluble proteins, the supernatant was stored at -20°C, while the pellets were resuspended in
10 2ml buffer C (6M guanidine hydrochloride, 100mM phosphate buffer, 10mM Tris-HCl, pH 7.5) and treated in a homogenizer for 10 cycles. The product was centrifuged at 13000rpm for 40 minutes.

Supernatants were collected and mixed with 150μl Ni²⁺-resin (Pharmacia) (previously washed with either buffer A or buffer B, as appropriate) and incubated at room temperature with gentle agitation
15 for 30 minutes. The sample was centrifuged at 700g for 5 minutes at 4°C. The resin was washed twice with 10ml buffer A or B for 10 minutes, resuspended in 1ml buffer A or B and loaded on a disposable column. The resin was washed at either (i) 4°C with 2ml cold buffer A or (ii) room temperature with 2ml buffer B, until the flow-through reached OD₂₈₀ of 0.02-0.06.

The resin was washed with either (i) 2ml cold 20mM imidazole buffer (300mM NaCl, 50mM
20 phosphate buffer, 20mM imidazole, pH 8) or (ii) buffer D (urea 8M, 10mM Tris-HCl, 100mM phosphate buffer, pH 6.3) until the flow-through reached the O.D₂₈₀ of 0.02-0.06. The His-fusion protein was eluted by addition of 700μl of either (i) cold elution buffer A (300mM NaCl, 50mM phosphate buffer, 250mM imidazole, pH 8) or (ii) elution buffer B (urea 8M, 10mM Tris-HCl, 100mM phosphate buffer, pH 4.5) and fractions collected until the O.D₂₈₀ was 0.1. 21μl of each
25 fraction were loaded on a 12% SDS gel.

K) His-fusion proteins renaturation

10% glycerol was added to the denatured proteins. The proteins were then diluted to 20μg/ml using dialysis buffer I (10% glycerol, 0.5M arginine, 50mM phosphate buffer, 5mM reduced glutathione, 0.5mM oxidised glutathione, 2M urea, pH 8.8) and dialysed against the same buffer at 4°C for 12-

14 hours. The protein was further dialysed against dialysis buffer II (10% glycerol, 0.5M arginine, 50mM phosphate buffer, 5mM reduced glutathione, 0.5mM oxidised glutathione, pH 8.8) for 12-14 hours at 4°C. Protein concentration was evaluated using the formula:

$$\text{Protein (mg/ml)} = (1.55 \times \text{OD}_{280}) - (0.76 \times \text{OD}_{260})$$

5 L) His-fusion large-scale purification (ORFs 111-129)

500ml of bacterial cultures were induced and the fusion proteins were obtained soluble in buffer M1, M2 or M3 using the procedure described above. The crude extract of the bacteria was loaded onto a Ni-NTA superflow column (Quiagen) equilibrated with buffer M1, M2 or M3 depending on the solubilization buffer of the fusion proteins. Unbound material was eluted by washing the column with the same buffer. The specific protein was eluted with the corresponding buffer containing 500mM imidazole and dialysed against the corresponding buffer without imidazole. After each run the columns were sanitized by washing with at least two column volumes of 0.5 M sodium hydroxide and reequilibrated before the next use.

M) Mice immunisations

20µg of each purified protein were used to immunise mice intraperitoneally. In the case of ORFs 2, 4, 15, 22, 27, 28, 37, 76, 89 and 97, Balb-C mice were immunised with Al(OH)₃ as adjuvant on days 1, 21 and 42, and immune response was monitored in samples taken on day 56. For ORFs 44, 106 and 132, CD1 mice were immunised using the same protocol. For ORFs 25 and 40, CD1 mice were immunised using Freund's adjuvant, rather than Al(OH)₃, and the same immunisation protocol was used, except that the immune response was measured on day 42, rather than 56. Similarly, for ORFs 23, 32, 38 and 79, CD1 mice were immunised with Freund's adjuvant, but the immune response was measured on day 49.

N) ELISA assay (sera analysis)

The acapsulated MenB M7 strain was plated on chocolate agar plates and incubated overnight at 37°C. Bacterial colonies were collected from the agar plates using a sterile dracon swab and inoculated into 7ml of Mueller-Hinton Broth (Difco) containing 0.25% Glucose. Bacterial growth was monitored every 30 minutes by following OD₆₂₀. The bacteria were let to grow until the OD reached the value of 0.3-0.4. The culture was centrifuged for 10 minutes at 10000rpm. The

supernatant was discarded and bacteria were washed once with PBS, resuspended in PBS containing 0.025% formaldehyde, and incubated for 2 hours at room temperature and then overnight at 4°C with stirring. 100µl bacterial cells were added to each well of a 96 well Greiner plate and incubated overnight at 4°C. The wells were then washed three times with PBT washing buffer (0.1% Tween-20 in PBS). 200µl of saturation buffer (2.7% Polyvinylpyrrolidone 10 in water) was added to each well and the plates incubated for 2 hours at 37°C. Wells were washed three times with PBT. 200µl of diluted sera (Dilution buffer: 1% BSA, 0.1% Tween-20, 0.1% NaN₃ in PBS) were added to each well and the plates incubated for 90 minutes at 37°C. Wells were washed three times with PBT. 100µl of HRP-conjugated rabbit anti-mouse (Dako) serum diluted 1:2000 in dilution buffer were added to each well and the plates were incubated for 90 minutes at 37°C. Wells were washed three times with PBT buffer. 100µl of substrate buffer for HRP (25ml of citrate buffer pH5, 10mg of O-phenildiamine and 10µl of H₂O) were added to each well and the plates were left at room temperature for 20 minutes. 100µl H₂SO₄ was added to each well and OD₄₉₀ was followed. The ELISA was considered positive when OD₄₉₀ was 2.5 times the respective pre-immune sera.

O) FACScan bacteria Binding Assay procedure.

The acapsulated MenB M7 strain was plated on chocolate agar plates and incubated overnight at 37°C. Bacterial colonies were collected from the agar plates using a sterile dracon swab and inoculated into 4 tubes containing 8ml each Mueller-Hinton Broth (Difco) containing 0.25% glucose. Bacterial growth was monitored every 30 minutes by following OD₆₂₀. The bacteria were let to grow until the OD reached the value of 0.35-0.5. The culture was centrifuged for 10 minutes at 4000rpm. The supernatant was discarded and the pellet was resuspended in blocking buffer (1% BSA, 0.4% NaN₃) and centrifuged for 5 minutes at 4000rpm. Cells were resuspended in blocking buffer to reach OD₆₂₀ of 0.07. 100µl bacterial cells were added to each well of a Costar 96 well plate. 100µl of diluted (1:200) sera (in blocking buffer) were added to each well and plates incubated for 2 hours at 4°C. Cells were centrifuged for 5 minutes at 4000rpm, the supernatant aspirated and cells washed by addition of 200µl/well of blocking buffer in each well. 100µl of R-Phicoerytrin conjugated F(ab)₂ goat anti-mouse, diluted 1:100, was added to each well and plates incubated for 1 hour at 4°C. Cells were spun down by centrifugation at 4000rpm for 5 minutes and washed by addition of 200µl/well of blocking buffer. The supernatant was aspirated and cells

resuspended in 200 μ l/well of PBS, 0.25% formaldehyde. Samples were transferred to FACScan tubes and read. The condition for FACScan setting were: FL1 on, FL2 and FL3 off; FSC-H threshold:92; FSC PMT Voltage: E 02; SSC PMT: 474; Amp. Gains 7.1; FL-2 PMT: 539; compensation values: 0.

5 P) OMV preparations

Bacteria were grown overnight on 5 GC plates, harvested with a loop and resuspended in 10 ml 20mM Tris-HCl. Heat inactivation was performed at 56°C for 30 minutes and the bacteria disrupted by sonication for 10 minutes on ice (50% duty cycle, 50% output). Unbroken cells were removed by centrifugation at 5000g for 10 minutes and the total cell envelope fraction recovered by centrifugation at 50000g at 4°C for 75 minutes. To extract cytoplasmic membrane proteins from the crude outer membranes, the whole fraction was resuspended in 2% sarkosyl (Sigma) and incubated at room temperature for 20 minutes. The suspension was centrifuged at 10000g for 10 minutes to remove aggregates, and the supernatant further ultracentrifuged at 50000g for 75 minutes to pellet the outer membranes. The outer membranes were resuspended in 10mM Tris-HCl, pH8 and the protein concentration measured by the Bio-Rad Protein assay, using BSA as a standard.

Q) Whole Extracts preparation

Bacteria were grown overnight on a GC plate, harvested with a loop and resuspended in 1ml of 20mM Tris-HCl. Heat inactivation was performed at 56°C for 30 minutes.

20 R) Western blotting

Purified proteins (500ng/lane), outer membrane vesicles (5 μ g) and total cell extracts (25 μ g) derived from MenB strain 2996 were loaded on 15% SDS-PAGE and transferred to a nitrocellulose membrane. The transfer was performed for 2 hours at 150mA at 4°C, in transferring buffer (0.3 % Tris base, 1.44 % glycine, 20% methanol). The membrane was saturated by overnight incubation at 4°C in saturation buffer (10% skimmed milk, 0.1% Triton X100 in PBS). The membrane was washed twice with washing buffer (3% skimmed milk, 0.1% Triton X100 in PBS) and incubated for 2 hours at 37°C with mice sera diluted 1:200 in washing buffer. The membrane was washed twice and incubated for 90 minutes with a 1:2000 dilution of horseradish peroxidase labelled anti-

mouse Ig. The membrane was washed twice with 0.1% Triton X100 in PBS and developed with the Opti-4CN Substrate Kit (Bio-Rad). The reaction was stopped by adding water.

S) Bactericidal assay

MC58 strain was grown overnight at 37°C on chocolate agar plates. 5-7 colonies were collected and used to inoculate 7ml Mueller-Hinton broth. The suspension was incubated at 37°C on a nutator and let to grow until OD₆₂₀ was 0.5-0.8. The culture was aliquoted into sterile 1.5ml Eppendorf tubes and centrifuged for 20 minutes at maximum speed in a microfuge. The pellet was washed once in Gey's buffer (Gibco) and resuspended in the same buffer to an OD₆₂₀ of 0.5, diluted 1:20000 in Gey's buffer and stored at 25°C.

50µl of Gey's buffer/1% BSA was added to each well of a 96-well tissue culture plate. 25µl of diluted mice sera (1:100 in Gey's buffer/0.2% BSA) were added to each well and the plate incubated at 4°C. 25µl of the previously described bacterial suspension were added to each well. 25µl of either heat-inactivated (56°C waterbath for 30 minutes) or normal baby rabbit complement were added to each well. Immediately after the addition of the baby rabbit complement, 22µl of each sample/well were plated on Mueller-Hinton agar plates (time 0). The 96-well plate was incubated for 1 hour at 37°C with rotation and then 22µl of each sample/well were plated on Mueller-Hinton agar plates (time 1). After overnight incubation the colonies corresponding to time 0 and time 1 hour were counted.

Table II gives a summary of the cloning, expression and purification results.

TABLE II – Summary of cloning, expression and purification

ORF	PCR/cloning	His-fusion expression	GST-fusion expression	Purification
orf 1	+	+	+	His-fusion
orf 2	+	+	+	GST-fusion
orf 2.1	+	n.d.	+	GST-fusion
orf 4	+	+	+	His-fusion
orf 5	+	n.d.	+	GST-fusion
orf 6	+	+	+	GST-fusion
orf 7	+	+	+	GST-fusion
orf 8	+	n.d.	n.d.	

orf 9	+	+	+	GST-fusion
orf 10	+	n.d.	n.d.	
orf 11	+	n.d.	n.d.	
orf 13	+	n.d.	+	GST-fusion
orf 15	+	+	+	GST-fusion
orf 17	+	n.d.	n.d.	
orf 18	+	n.d.	n.d.	
orf 19	+	n.d.	n.d.	
orf 20	+	n.d.	n.d.	
orf 22	+	+	+	GST-fusion
orf 23	+	+	+	His-fusion
orf 24	+	n.d.	n.d.	
orf 25	+	+	+	His-fusion
orf 26	+	n.d.	n.d.	
orf 27	+	+	+	GST-fusion
orf 28	+	+	+	GST-fusion
orf 29	+	n.d.	n.d.	
orf 32	+	+	+	His-fusion
orf 33	+	n.d.	n.d.	
orf 35	+	n.d.	n.d.	
orf 37	+	+	+	GST-fusion
orf 58	+	n.d.	n.d.	
orf 65	+	n.d.	n.d.	
orf 66	+	n.d.	n.d.	
orf 72	+	+	n.d.	His-fusion
orf 73	+	n.d.	+	n.d.
orf 75	+	n.d.	n.d.	
orf 76	+	+	n.d.	His-fusion
orf 79	+	+	n.d.	His-fusion
orf 83	+	n.d.	+	n.d.
orf 84	+	n.d.	n.d.	
orf 85	+	n.d.	+	GST-fusion
orf 89	+	n.d.	+	GST-fusion
orf 97	+	+	+	GST-fusion
orf 98	+	n.d.	n.d.	
orf 100	+	n.d.	n.d.	
orf 101	+	n.d.	n.d.	
orf 102	+	n.d.	n.d.	
orf 103	+	n.d.	n.d.	
orf 104	+	n.d.	n.d.	
orf 105	+	n.d.	n.d.	
orf 106	+	+	+	His-fusion
orf 109	+	n.d.	n.d.	
orf 110	+	n.d.	n.d.	

09303516-043009

orf 111	+	+	n.d.	His-fusion
orf 113	+	+	n.d.	His-fusion
orf 115	n.d.	n.d.	n.d.	
orf 119	+	+	n.d.	His-fusion
orf 120	+	+	n.d.	His-fusion
orf 121	+	n.d.	n.d.	
orf 122	+	+	n.d.	His-fusion
orf 125	+	+	n.d.	His-fusion
orf 126	+	+	n.d.	His-fusion
orf 127	+	+	n.d.	His-fusion
orf 128	+	n.d.	n.d.	
orf 129	+	+	n.d.	His-fusion
orf 130	+	n.d.	n.d.	
orf 131	+	+	+	n.d.
orf 132	+	+	+	His-fusion
orf 133	+	n.d.	+	GST-fusion
orf 134	+	n.d.	n.d.	
orf 135	+	n.d.	n.d.	
orf 136	+	n.d.	n.d.	
orf 137	+	n.d.	+	GST-fusion
orf 138	+	n.d.	+	GST-fusion
orf 139	+	n.d.	n.d.	
orf 140	+	n.d.	n.d.	
orf 141	+	n.d.	n.d.	
orf 142	+	n.d.	n.d.	
orf 143	+	n.d.	n.d.	
orf 144	+	n.d.	+	n.d.
orf 147	+	n.d.	n.d.	

Example 1

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 1>:

```

5      1  ATGAACACAGA  CAGTCAA.AT  GCTTGCCGCC  GCCGTGATTG  CTTGGGCTTT
      51  GAACCGACCG  GTGTGGNCGG  ATGACGTATC  GGATTTTCGG  GAAACCTTGC
     101  A.GCGGCAGC  ACAGGGAAAT  GCAGCAGCCC  AATACATTTT  GGGCGCAATG
     151  TAT.TACAAA  GGACGCGCGT  GCGCGGGGAT  GATGCTGAAG  CGGTGAGATG
     201  GTATCGGCAG  CCGCGCGAAC  AGGGSTTAGC  CCAAGCCCAA  TACAATTTGG
     251  GCTGGATGTA  TGCCAAACGG  CCGCG.GTGC  GCCAAGATGA  TACCGAAGCG
     301  GTCAGATGGT  ATCGGCAGGC  GGCAGCGCAG  GGGGTTGTCC  AAGCCCAATG
     351  CAATTTGGGC  GTGATATATG  CGAAGGACG  TGGAGTGCGC  CAAGACGATG
     401  TCGAAGCGGT  CAGATGGTTT  CGGCAGGCGG  CAGCGCAGGG  GGTAGCCCAA
     451  GCCCAAAACA  ATTTGGGCGT  GATGTATGCC  GAAAGANGCG  CGGTGCGCCA
    15      501  AGACCG...

```

This corresponds to the amino acid sequence <SEQ ID 2; ORF37>:

```

1  MKQTVMXLLAA ALIALGLNRP VWKDDVSDFR ENLXAAQGN AAAQYNLGM
51  YXQTRVRERD DAEAVRWYRQ PAEGQLAQAG YNLGWMYANG RXVRQDDTEA
101 VQNYRQAAAGG GVVQAQYNLG VIYAEGRGVR QDDVEAVRVF RQAAAQGVAG
151 ARNNRLGVMA ERXVRVRD. . .

```

5 Further work revealed the complete nucleotide sequence <SEQ ID 3>:

	1	ATGACACAGA	CAGTCAATATG	GCTTCGCCGC	GCCTTGATTG	CCTTGGGGTT
	51	GACCCGAGCG	GTGTCGGCGG	GTGATCATTC	GGGATTTTCG	GAAACATTGC
	101	AGGCGCGAGC	ACGAGGAAAT	CGACGCGCAAT	AATACAAATT	GGGCGCAATT
10	151	TATTACAAGG	GAGCGGGCGT	CGCGCGGGAT	GATGCTGAAG	CGCTCAGATG
	201	GTCGTCGAGC	CGCGCGGAAAT	AGGGGTTAGC	CCGAGCCCAA	TACAATTTGG
	251	CTGTTAGTGA	TGCCAACGGG	CGCGGCTGCG	SCCAAGATATG	TACCGAAGCG
	301	GTGAGATGTG	GTGACGAGCG	GGCAGCGGAC	GGGGTTGTCC	AAGCCCAAAT
	351	CAATTTTGGC	ATGATATATG	CCGAGGACAG	TGGAGTGGCG	CAAGACAGAT
	401	TCCGAGCGGT	CAGATGGTTT	CGGACGCGGG	CAGCGCAGGG	GGTAGCCCAA
15	451	GCGCAAAACA	ATTTGGGGCT	GATTGTATCG	GATGTCGGCA	CGCTGGCCCA
	501	AGACCGCGCC	CTTGCACAA	AATCGTTTGG	CACGGCTTGT	CAAAACGGAG
	551	ACCAGACAGG	CTGCACAAAT	GACCAACCGC	TGAGGCGGGG	TTATTGTA

This corresponds to the amino acid sequence <SEQ ID 4; ORF37-1>:

20

1	MKQTVKWLAA	ALIALGLNRA	VWADDVSDFR	ENLQAAQGN	AAQYNLGM
51	YYKGRGVRRD	DAEAVRWYRQ	AAEQGLAQ	YNLGMWYANG	RGVRQDDTEA
101	VWRVYRQAAQ	GVVQAYNLG	VIYAEGRGVR	QDDVEAVRWF	RQAAQGVVAQ
151	AOPLLGVMYA	ERRGVRODRA	LAQEWFGKAC	QNGDVGCDNF	DORLKAGY*

Further work identified the corresponding gene in strain A of *N.meningitidis* <SEO ID 5>:

25

1	ATGAAACAGA	CAGTCAAATG	GCTTCGCCGC	GGCCTGATTT	CCTTGGGGCTT
51	AACCAACAGC	GCTTGCGCGG	ATGACGTATC	GATATTTCGG	GAAACATTCG
101	AGCGCGCGAC	ACAGGGAAAT	CGACGACGCC	AAACCAATTT	GGGGCTGATG
151	TATGCCGAAA	CGCGCGCGCT	CGCCGACAA	CGCGCCCTT	CACAGAAATG
201	GCTTGGCAAG	GCTTGTCAAA	ACGGATACCA	AGACAGCTGC	GACAATGACC
251	AACGCGTTAA	AGCGGTTAT	TGA		

30 This encodes a protein having amino acid sequence <SEQ ID 6; ORF37a>:

1 MKQTVKWLAA ALIALGLNQA VWADDVSDFR ENLQAAAQGN AAAQNNLGVM
51 YAERRGVROD RALAQEWLGK ACQNGYQDSC DNDORLKAGY *

The originally-identified partial strain B sequence (ORF37) shows 68.0% identity over a 75aa overlap with ORF37a;

40

orf37.pep	10	20	30	40	50	60
	<u>MKQTVKMLAALIALGLNRPVFWDDVSDFRENLQAAAGNAAQYNLGMAYGTRQVTRVR</u>					
	: : :					
orf37a	<u>MKQTVKWLAAALIALGLNQAVWDDVSDFRENLQAAAGNAAQNNLGMVYAEERRGQY</u>					
	10 20 30 40 50 60					

45

orf37.pep	70	80	90	100	110	120
	<u>DAEAVRWYRQPAEQGLGAQYNLGMWYANGRXVRQDDTEAVRWYRQAAAQGVVQAQYNL</u>					
	: : : : : : : : : : : : : : :					
orf37a	<u>RALQGEWLKGAQGGYQDSCDNDRLKAGYX</u>					
	70 80 90					

Further work identified the corresponding gene in *N.gonorrhoeae* <SEQ ID 7>:

50

1	TGAACAAACAG	CAGTCAAAATG	GCTTCCGCCCG	GGCTTGATTG	GCTTGGGCCTT
51	GATCCACACCG	GTACGGTGGCG	GTAGCGTATTC	GTAATTTGCG	GAACAACCTTGC
101	AGggcgcaCAG	ACaagCAAAAT	CCAGCAGGCC	AAATTCATGT	GGCGCTCAGT
151	TATGAAAAATG	CAACAAGAGT	GCTTCAAGAT	TATGTACAGG	CATGTCAGTGC
201	GTATGCACAG	GCTTTCAGAAC	AGGGGGATGCT	CCAGACCCAA	TACAATTTTGG
251	CGTTGATGTA	TTACGATGGA	GCGGCGCTGC	CCCAAGACCT	TGGCGCTCGT
301	CAACAATGGC	TTGGCAGAGC	TTGTCAAAAC	GGAGACCAA	ACAGCTGCGA
351	CAATGACCAA	CGGCTGAAGG	CGGCTTATTA	Δ	

55

This encodes a protein having amino acid sequence <SEQ ID 8; ORF37ng>:

```

1  MKQTVKWLAA ALIALGLNQA VWAGDVSDFR ENLQAAEQGN AAQFNLGVM
51 YENGQGVROD YVQAVQWYRK ASEQGDAAQ YNLGLMYDG RGRVQDLALA
101 QQWLKACQN GDQNSCNDQ RLKAGY*

```

- 5 The originally-identified partial strain B sequence (ORF37) shows 64.9% identity over a 111aa overlap with ORF37ng:

```

orf37.pep  MKQTVXMLAAALIALGLNRPVWDDVSDFRENLXAAAQGNAAAQYNLGAMYXQRTVRRD 60
10 orf37ng  MKQTVKWLAAALIALGLNQA VVWAGDVSDFRENLQAAEQGNAAAQYNLGVMYENGQGVROD 60
orf37.pep  DAEAVRWYRQPAEQGLAQAYNLGWMYANGRXVRQDDTEAVRWYRQAAAQGVVQAQYNLG 120
orf37ng  YVQAVQWYRKASEQGDAAQAYNLGLMYDGRGVRODLALAQWLKACQNGDQNSCNDQ 120
15 orf37.pep  VIYAEGRGVRODDVEAVRWFRQAAAQGVQAQNNLGVMYAERXVRQD 168
orf37ng  RLKAGY 126

```

The complete strain B sequence (ORF37-1) and ORF37ng show 51.5% identity in 198 aa overlap:

```

20 orf37-1.pep  10      20      30      40      50      60
orf37ng  MKQTVKWLAAALIALGLNRAVWADDVSDFRENLQAAEQGNAAAQYNLGAMYXKGRVRRD
100 orf37ng  MKQTVKWLAAALIALGLNQA VVWAGDVSDFRENLQAAEQGNAAAQYNLGVMYENGQGVROD
orf37-1.pep  70      80      90      100     110     120
25 orf37ng  DAEAVRWYRQAAEQGLAQAYNLGWMYANGRGVRQDDTEAVRWYRQAAAQGVVQAQYNLG
orf37ng  YVQAVQWYRKASEQGDAAQAYNLGLMYDGRGVROD-----
30 orf37-1.pep  130     140     150     160     170     180
orf37ng  VIYAEGRGVRODDVEAVRWFRQAAAQGVQAQNNLGVMYAERRGVRQDRALAQEWFGKAC
35 orf37ng  -----LALAQWLKAC
100
orf37-1.pep  190     199
40 orf37ng  QNGDQDGCNDQRLKAGYX
110     120
orf37ng  QNGDQNSCNDQRLKAGYX
110     120

```

Computer analysis of these amino acid sequences indicates a putative leader sequence, and it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

- 45 ORF37-1 (11kDa) was cloned in pET and pGEX vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 1A shows the results of affinity purification of the GST-fusion protein, and Figure 1B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result), FACS analysis (Figure 1C), and a

bactericidal assay (Figure 1D). These experiments confirm that ORF37-1 is a surface-exposed protein, and that it is a useful immunogen.

Figure 1E shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF37-1.

Example 2

- 5 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 9>:

TTCTGGCGA	TTGGGCGGGT	TTGAAGTTC	ATGCGCCCGT	CAAAATCGCG
GGCTGATTTG	TCGGGCGCGT	OGGCGGTC	GGACTTGACC	CGAAATTCCTA
TCRAGGCGCG	TCGGCGCCTC	ACTTTGAGCG	CAAGTATFCG	TTTCRAGGAGC
ACGTTTTCGG	GCAAAATCTG	ACCTTCGGAC	TTTTCGGCGA	CGCATCATCTC
GGGCTGCGAC	AGGGGCGCGA	CACGGAAAGC	CTTCGTGTCG	CGACACACAT
CTCCGTAGAC	AGTTTCTCGA	TGGTTCATGA	AAACCTTCAT	GGCAAAATCA
TGACGAGATT	TGCCGAGAAA	AATGCGGACG	GCGCAATGC	GGAAAAAGCC
CGCGAATBA				

This corresponds to the amino acid sequence <SEO ID 10>:

```

1  FGDIGGLKVN  APVKSAGVLV  GRVGAIGLDP  KSYQARVRLD  LDGKYQFSSD
51  VSAQILTSGL  LGEQYIGLQQ  GGDTENLAAG  DTISVTSSAM  VLENLIGKFM
101 TSFAEKNADG  GNAEKAEE*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a hypothetical *H.influenzae* protein (ybrd.haein; accession number p45029)

- 20 SEQ ID 9 and ybrd.haein show 48.4% aa identity in 122 aa overlap:

20 30 40 50 60 70
 yrbd.h LGIGALVFLGRVANVQGFATKSYTVTATFDNIGGLKVRAPLKIIGVVIGRVSAITLDE
 N.m FGDIGGLKVNAPVKSAGVLVGRVGAIGLDP
 10 20 30

80 90 100 110 120 130
 yrbd.h KSYLPKVSIAINQEYNEIPENSSLSIKTSGLLGEQYIALTMGFDDGTAMLKNGSIQDT
 ||| ::::: | ::::: | ||||| ::::: | ||| ::::: |
 N.m KSYQARVRLDLGKY-QFSDDVSAQILTSGLLGEQYIGLQG---GDTENLRAAGDTISVT
 40 50 60 70 80

```

      140          150          160
yrbd.h TSAMVLEDLIGQFL--YGSKKSDGNEKSESTEQ
       :|:||||:||||:FM: ::::|:: :::::|
N.m    SSAMVLENLICGFMTSFAEKNADGGNAEKAAEX
        90         100         110         120

```

Homology with a predicted ORF from *N.gonorrhoeae*

- 40 SEQ ID 9 shows 99.2% identity over a 118aa overlap with a predicted ORF from *N. gonorrhoeae*:

yrbd
N.m

20 30 40 50 60 70
GAAVAFLAFRVAGGAFFGSQKTYAVYADFPGDIGGLKNVAPVKSGVLGRVGAIGLDP
| | | | |
FGDIGGLKNVAPVKSGVLGRVGAIGLDP
10 20 30

80 90 100 110 120 130
KSYQARVRLDLDGKYQFSSDVSAQILTSLGLGEQVIGLQGSDTENLAAGDTISVTSSAM
| | | | |
KSYQARVRLDLDGKYQFSSDVSAQILTSLGLGEQVIGLQGSDTENLAAGDTISVTSSAM
40 50 60 70 80 90

		140	150	160
yrbd		VLENLIGKFM	TSFAEKNAEGGNAEKAAEX	
5	N.m	VLENLIGKFM	TSFAEKNAEGGNAEKAAEX	
		100	110	120

The complete yrbd *H.influenzae* sequence has a leader sequence and it is expected that the full-length homologous *N.meningitidis* protein will also have one. This suggests that it is either a membrane protein, a secreted protein, or a surface protein and that the protein, or one of its epitopes, could be a useful antigen for vaccines or diagnostics.

Example 3

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 11>:

```

1  ..ATTTTGATAT ACCTCATCG CAAGAATCTA GGTTCGCCG TCTTCTCTT
51 TCAGGAACGC CCGGAAAGG ACGGAAAAC TTTTAAATG GTCAAAATTC
101 GTTCCATCGC CGACGGCTTG TATTGACAG CGATTCCGCT GCCCGACGGA
151 GAACGCGTGA CACGCTTCG CAAAAAACTG CGTGGCGCA GTWTGGACGA
201 ACTGCTGAA TTATGGAATA TCTTAAAGG CGAGATGAGC CTGGTCGGCC
251 CCCGCCCGCT GCTGATGCAA TATCTGCGCG TGTAAGCAA CTTCGAAAC
301 CCGCGCCACG AATGAAACC CGGCATTACC GGCTGGGCGC AGGTCAACGG
351 CGCCACGCG CTTCGTGGG ACGAAAATT CGCCTCGGAT GTTTGGTATA
401 TCGACCACTT CAGCCTGTGC CTCGACATCA AATCCTACT GCTGACGGTT
451 AAAAAGTAT TAATCAAGGA AGGATTTCG GCACAGGGG ACACAACAT
501 GCCCCTTTC ACAGGAAAAC GCAACTCGC CGCTGTCGT CGCGGCGGAC
551 ACGGAAAGT CGTTCGCGAC CTTGCCGCG CACTCGGCCG GTACAGGGAA
601 ATCGTTTTTC TGGACGACG GCGACAAGG AGCGTCAACG GCTTTTCGT
651 CATCGGACG ACGCTGCTGC TTGAAAACAG TTTATCGCCC GAACAATACG
701 ACGTCGCGCT CGCGCTCGG AACAAACCGA TCCGCCGCCA ATTCGCCGAA
751 AAAGCCGCCG CGCTCGGCTT CGCCCTGCC GATACTGGTT ATCCGCCGAC
801 GACCGTCTCG CCTCTGCAA CAGTCGGACA AGGCAGCGTC GTTATGGCGA
851 AAGCGGTG.

```

This corresponds to the amino acid sequence <SEQ ID 12; ORF3>:

```

1  ..ILIYLRKNL GSPVFFOER PGKDGKPFKM VKFRSMRDL YSDGIPLPDG
51 ERITPFKKL RAASXDELPE LWNILKEMS LVGFRLIMQ YLFLYDNFQN
101 RRHEMKFPIG WAGCVNGRNA LSNDEKPCD VVYDIHSLC LDKILLITV
151 KKVLIEKGIS AQGXMPFPF TGRKRLAVVG AGGHKLVAVD LAALGRVRE
201 IVFLDDRAGG SVNGFSVIG TLLLENSLP EYDVAVAVG NNRIRQIAE
251 KAAALGFALP VLVHPDATVS PSATVGQGSV VMAKAV..

```

Further sequence analysis revealed the complete nucleotide sequence <SEQ ID 13>:

```

1  ATGAGTAAT TCTTCAAAC CTTGTTTGAC ATTGTTGCT CCGCTCGGG
51 ACTGATTTTC CTCTGCCAG TATTTTTGT TTTGATATAC CTCATCCGCA
101 AGAATCTAGG TTGCGCCGCT TTCTTCTTC AGGAACGCC CGGAAAGGAC
151 GGAAACCTTT TTAATATGGT CAATATCCGT TCCATGCGCG ACGCGTTTGA
201 TTCAGACGGC ATTCCGCTGC CGACGGAGA ACGCTGAGC CCGTTCGGCA
251 AAAAAGTCTG TCGCGCAAT TTGGACGAA TGGCTGAAT ATGGAATATC
301 TTAAGAAGCG AGATGACCT GTTCGCGCC CGCCGCTGC TGATGCAATA
351 TCTGCCGCTG TAGACAAC TCCAAACCG CCACGCGAA ATGAACCGG
401 GCATTACCGG CTGGCGCAG GTCAACGGCG GCAACGGCT TCTGTCGAC
451 GAAAAATTCG CTTGCGATGT TTGTTATATC GAACACTTCA GCCTGTGCTT
501 CGACATCAAA ATCTACTGCT TGACGGTTAA AAAAGTATTA ATCAAGGAAG
551 GGATTTCCGC ACAGGGCGAA GCCACCATGC CCCCCTTCC AGGAAACCGC
601 AAATCTCGCG TCGTCCGTGC GGGCGGACAC GAAAAGTCTG TTGCGACCT
651 TCGCGCGCCA CTCGCGCGGT ACAGGGAAT CTTTTCTG TCGGACGAC GGTGCTGCTT
701 CACAAGCGAG CGTCAACGGC TTTTCGTGA TCGGACGAC GGTGCTGCTT
751 GAAAACGAGT TATCGCCGCA ARAATACGAC GTGCGCGTGC CGTTCGGCAA
801 CAACGCATC CGCGCCCAA TCGCGGAAA AGCGCGCGC CTCGCTTTG
851 CCTGCGCGT TCTGTTTCA TCGGACGCG CCGTCTCGC TTTGCAACA

```

5
 901 GTCGGACAAG GCAGCGTCGT TATGGCGAAA GCCTCGTAC AGGCAGGCAG
 951 CGTATTGAAA GACGGCGTGA TTGTGAACAC TGCCGCGACC GTGCATCACG
 1001 ACTGCTGCT TAACGCTTTC GTCCACATCA GCCCAGCGCG GCACCTGTCTG
 1051 GCACACACGC ATATCGGCGA AGAAAAGCTGG ATAGGACCGG CGCGCTGCACG
 1101 CGCGCAGCAG ATCCGTATCG GCACGCGCGC AACCATTCGA CGCGGCGCAG
 1151 TCGTCGTACG CGACGCTTCA GACGCGATGA CCGTCCGCG CAATCCGCGA
 1201 AAGCCGCTGC CGCGCAAAA CCCGAGACC TCGACGCAT AA

This corresponds to the amino acid sequence <SEQ ID 14; ORF3-1>:

10
 1 MSKFFKRLFD IVASAGSLIF LSPVFLILIY LIRKNLGSFV FFFQERPGKD
 51 GKPFKVMKFR SMRDALDSGD IPLPDGERLT PFGKILRAAS LDELPELWNI
 101 LKGEMSLVGP RPLLMQYLPL YDNFQNRHE MKPGITGWAQ VNGRNLSDW
 151 EKFACDVWYI DHFSLCLDIK ILLTVKKVL IKEGISAQGE ATMPPFTGKR
 201 KLAUVGAGGH GKVVADLAAA LGRYREIVFL DORAQGSVNG FSVIGTLLLL
 15
 251 ENSLSPEQYD VAVAVGNNRI RRQIAEKAAL LGFALPVLVH PDATVSPSAT
 301 VGQGSVVMKAV AVVQAGSVLK DGVIVNTAAT VDHDCLLNAF VHISPGAHL
 351 GNTHIGEESS IGTGACSRQQ IRIGRSRTIG AGAVVVRDVS DGMTVAGNPA
 401 KPLPRKNPET STA*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

20 ORF3 shows 93.0% identity over a 286aa overlap with an ORF (ORF3a) from strain A of *N. meningitidis*:

25
 orf3.pep 10 20 30
 ILLIYLIRKNLGSVPVFFQERPGDKGKPFKVMKFR
 orf3a 10 20 30 40 50 60
 MSKFFKRLFDIVASAGSLIFSVFLIYLIRKNLGSVPVFFQERPGDKGKPFKVMKFR
 30
 orf3.pep 40 50 60 70 80 90
 SMRDGLYSDDGILPLDGERLTPFGKILRAASXDELPELWNLKGEMLSVGRPLLMQYLPL
 orf3a 40 50 60 70 80 90 100 110 120
 SMRDALDSGILLPLDGERLTPFGKILRAASLDELPELWNLKGEMLSVGRPLLMQYLPL
 35
 orf3.pep 100 110 120 130 140 150
 YDNFQNRHEMKPGITGWAQVNGRNLSDWDEKFACDVWYIDHFSCLDIKILLTVKKVL
 orf3a 100 110 120 130 140 150 160 170 180
 YDNFQNRHEMKPGITGWAQVNGRNLSDWDERFACDIWYIDHFSCLDIKILLTVKKVL
 40
 orf3.pep 160 170 180 190 200 210
 IKEGISAQGEATMPPTGKRKLAVVAGGHHKVVADLAAALGRYREIVFLDORAQGSVNG
 orf3a 160 170 180 190 200 210 220 230 240
 IKEGISAQGEATMPPTGKRKLAVVAGGHHKVVADLAAALTYGEIVFLDRVQGSVNG
 45
 orf3.pep 220 230 240 250 260 270
 FSVIGTLLLLNSLSPEQYDVAVGNNRIIRQIAEKAALGFALPVLVHPDATVSPSAT
 orf3a 220 230 240 250 260 270 280 290 300
 FSVIGTLLLLNSLSPEQYDVAVGNNRIIRQIAEKAALGFALPVLVHPDSTVSPSAT
 50
 orf3.pep 280
 VGQGSVVMKAV
 orf3a 280 290 300 310 320 330 340 350 360
 VGQGSVVMKAVVQADSVLKDGVIVNTAATVDHDCLLDFAVHISPGAHLGNTRIGEESS

The complete length ORF3a nucleotide sequence <SEQ ID 15> is:

60
 1 ATGAGTAAAT TCITCAAACG CCTGTTTGAC ATTGTTGCCT CCGCCTCGGG
 51 ACTGATTTC CTCTCGCCAG TATTTTGTAT TTGATATAC CTCATCCGCA
 101 AGAATCTGGG TTGCGCGTC TTCTCTTTC AGGAACGCC CGGAAGAGCA

151 GGAAACCTT TAAAAATGGT CAAATTCOGT TCCATGCACG ACGCGCTTGA
 201 TTCAGACGGC ATTCGTGCTGC CGACGGGAGA ACGCCTGACA CGTTCGGCA
 251 AAAAATCTGGC TGCGCGCAGT TTGACGAAAC TGCCCCGACT GTGGAACCTC
 351 CTCGAAGGCG ACATGAGCCT GGTGGGCGCC TGAATGCAATA TGATGCAATA
 401 GCATTACCGC CTGGGCGCAG GTCAACGGGC GCAACCGCCT TTCGTGGGAC
 451 GAACGCTTCG CATCGGACAT CTGATATATC GACCACTTCA GCCTGTGCCT
 501 CGACATCAAA ATCCTACTGC TGACGGTTAA AAAAGTATTA ATCAAGAAGAG
 551 GGAATTTCCGC ACAGGGCGAA GCCACATCG CCCTTTTCAC AGGAAAACGC
 601 AAATCTTGGC TCGTGGTGC GGGCGGACAC GGCAAGATCG TTGCGGAGCT
 651 TGCGCGCCGA CTGCGGCACAT ACGCGGAAAT GGTTTTCTG GACGACCGCG
 701 TCCAAGCGAG GTCAACGGC TTCCCGCTCA TCGGCACGAC GCTGCTGCTT
 751 GAAAAACAGTT TATCGCCGGA ACAATTTCAG ATCGCCGTGC CGTTCGGCAA
 801 CAACCGCATC CGCGCGCAAA TCGCCGAAAA AGCGCGCGCG CTGCGGCTTG
 851 CCCTGCCCGT CCTGATTCAT CGGACTCGA CGCTCTCGCC TTCTGCAACA
 901 GTCGGACAG GCGCGCTGCT TATGGCGAAA GCGCTCGTAC AGGCTGACAG
 951 CGTATTGAAA GACGCGGTAA TTGTGAACAC TGCGCGCAAC GTCGATCAOG
 1001 AITGCTGCTT TGATGCTTTC GTCCACATCA GCCCGGGCGC GCACCTGTGC
 1051 GCGAACGCG GTATCGCGGA AGAACACCTG ATGACGACAG CGCGTTCGAG
 1101 CGCCGACGAG ATCCGTATCG CAGCGCGCGC AACCATTTGA GCGGCGCGAG
 1151 TCGTCTGTCG CGACGTTTCA GACGGCATGA CCGTCCGCGG CAACCGCGCA
 1201 AAACCATTTG CAGGCAAAAA TACCGAGACC CTGCGTCTGT AA

This is predicted to encode a protein having amino acid sequence <SEQ ID 16>:

1 MSKFFKRLFD IVASASGLIF LSPVFLILYI LIRKNLGSPV FFFQERPGKD
 51 GKPFKMKVFR SMHDAALSDG ILLPDGERLT PFGKILRAAS LDELPELWNV
 101 LKGDMSLVGP RFLIMQYLPL YDNFQNRHLE MKPGITGWAQ VNGRNLASWD
 151 ERFACDIWYI DHFSLCLDIK ILLLTGVKVL IKEGISAQGE ATMPFPTGKR
 201 KLAUVGAGGH GKVVAEALAA LGTYGELIVL DDRVQGSVNG FVIGITLLLL
 251 ENSLSPEQFD IAVAVGNMRI RRQIAEKAAA LGFALFVLII PDSTVSPSAT
 301 VGQGGVVMK AVVQADSVLK DGLIVNTAAT VDHCLLDLDF VHISPGAHLS
 351 GNTRIGEESS IGTGACSRQQ IRIIGSRATIG AGAVVVRDVS DGMTVAGNPA
 401 KFLAGKNTET LRS*

Two transmembrane domains are underlined.

ORF3-1 shows 94.6% identity in 410 aa overlap with ORF3a:

orf3a.pep 10 20 30 40 50 60
 MSKFFKRLFDIVASASGLIFLSPVFLILYI LIRKNLGSPVFFQERPGKD GKPFKMKVFR
 orf3-1 MSKFFKRLFDIVASASGLIFLSPVFLILYI LIRKNLGSPVFFQERPGKD GKPFKMKVFR
 70 80 90 100 110 120
 orf3a.pep SMHDAALSDGILLPDGERLT PFGKILRAASLDELPELWNV LKGDMSLVGP RFLIMQYLPL
 orf3-1 SMHDAALSDGILLPDGERLT PFGKILRAASLDELPELWNV LKGDMSLVGP RFLIMQYLPL
 130 140 150 160 170 180
 orf3a.pep YDNFQNRHHEMKPGITGWAQVNGRNLASWDERFACDIWYIDHFS LCLDIKILLTVKKVL
 orf3-1 YDNFQNRHHEMKPGITGWAQVNGRNLASWDEKFACDVWYIDHFS LCLDIKILLTVKKVL
 190 200 210 220 230 240
 orf3a.pep IKEGISAQGEATMPPTGKRKLAVVGAGGHGKVVAEALAAALGTGTYGELV DDRVQGSVNG
 orf3-1 IKEGISAQGEATMPPTGKRKLAVVGAGGHGKVVADLAAALGRYREIVLDDRAQGSVNG
 250 260 270 280 290 300
 orf3a.pep FVIGITLLLENLSLSPQFDIAVAVGNMRI RRQIAEKAAAALGFALFVL IHPDSTVSPSAT
 orf3-1 FVIGITLLLENLSLSPQFDIAVAVGNMRI RRQIAEKAAAALGFALFVL IHPDSTVSPSAT

5

	310	320	330	340	350	360
orf3a.pep	VGQGVSVMAKAVVQADSLVLDKGVIVNTAAATVDHCDLDAFVHISPGAHLSGNTRIGEE	SW				
orf3-1	VGQGVSVMAKAVVQAGSLVLDKGVIVNTAAATVDHCDLLNAFVHISPGAHLSGNTHIGEE	SW				

10

	370	380	390	400	410
orf3a.pep	IGTGACSRQRIIRIGSRATIGAGAVVVRDSDGMTVAGNPAKPLAGNNTETLSRX				
orf3-1	IGTGACSRQRIIRIGSRATIGAGAVVVRDSDGMTVAGNPAKPLPRKNPETSTAX				

Homology with hypothetical protein encoded by *yvfc* gene (accession Z71928) of *B. subtilis*

15 ORF3 and YVFC proteins show 55% aa identity in 170 aa overlap (BLASTp):

	ORF3	3	YLIRKWLNSGVPTFFQERPGDKGDFPKMYKFSRMDGLYSDDGSLPLDGERLRTTPGKKLRA	62
			I++R+GSPVFQ R RQG KCFP +KFR+M+ S G LFD RLT G+K+R	
	yvfc	27	IAVVLRLKIGSPVFKQVFRGLHGKPFYLYKFTTMTDERDSKGNLLPDEVLRTKGRGLRK	86
20	ORF3	63	ASXDELPELWNILKMGESLVGPRFLLMQYPLPLDYNQNRNHEMKMGITGVAQVNGRNAS	122
			S DELP+L+N+LKG++SLVSGPRFLLM LYLPL Q R RHE+KPGITGVAQ+NGRNA+S	
	yvfc	87	LSDLELQQLNLWIKGDSLVLVSGPRFLLMDYLPLYTEQARRHEVKPGITGVAQINGRNAS	146
25	ORF3	123	NDEKDFADVNYIDWNSFLDCLXXLXXXXXXGISAQGXETMPFTG	172
			W++KF DVNY+D++S L D EGI T F T G	
	yvfc	147	WEKKFELADVNYIDWNSFLDCLXXLXXLTKRVKLVSGEIQQTNHVHTAERTG	196

Homology with a predicted ORF from *N.gonorrhoeae*

ORF3 shows 86.3% identity over a 286aa overlap with a predicted ORF (ORF3.ng) from *N.*

30 *gonorrhoeae*:

	orf3 <u>LLVYLIRKNL</u> SGSPVFFQERPGKDGKFKPMVKFR	34
	orf3ng	MSKAVKRLFDIIASAS <u>GLIVLSPVFLV</u> LVYLIRKNKGS PVFFIRERPGKDGKFKPMVKFR	60
35	orf3	SMRDGLYS DGI PLP DGERLT PFGKRLRAASXDELPELWNILKEMSLVGRPILMQYLP	90
	orf3ng	SMRDALSDGI PLP DLSERITDPFGKRLRALSDELPELWNVVKGEMSLVGRPILMQYLP	120
40	orf3	YDNFQNRHHEMKGPTIGWAQVNGNRLNSWDEKFCADVWYIIDHFSCLCIDILLTVKKVL	150
	orf3ng	YNKFQNRHHEMKGPTIGWAQVNGNRLNSWDEKFS CDVWYTDNFSFWLDMKILFLT VKKVL	180
45	orf3	IKEGISAQGEXTMPFFTGKRKLAVWGAGGCHKVVAADLAALGRYREIVFLDRAQGSVNG	210
	orf3ng	IKEGISAQGEATMPFFAGNRKLAVIGAGGCHKVVAE LAALALTYGEIVFLDRTQGSVNG	240
50	orf3	FSVIGITLLLENSLSPEQYDVAVAVGNNRIRQIAEKAALSFALPVLVHPDATVSPSAT	270
	orf3ng	FPVIGITLLLENSLSPEQPDITAVAVGNNRIRQITENAAALGFKLPVLHPDATVSPSAI	300
	orf3	VGQGSVVMKAV	286
	orf3ng	IGQGSVVMKAVVQAGSVLKGDIVNTAATVHDHCLLDVFIHSPGAHLSGNTRIGEESR	360

The complete length ORF3ng nucleotide sequence <SEQ ID 17> is:

55

1	ATGAGTAAAG	CGGTCAAAAG	CCTGTTGCAG	ATCATCGCAT	CCGCATCGGG
51	CTGCTATTGTC	CTGTGCGCCG	GTGTTTTGGT	TTTAATATAC	CTCATCGGCA
101	AAAACTTAGG	TTCCGCCCTG	TTCTTTCatC	GGGAACGCCc	cgAAAGCGa
151	ggaaaacCTT	TTAAATATGT	CAAAATTCGT	TCGATcgcg	acgcgcttGA
201	TTACAGACCG	ATTCCGCTGT	CGGATGAGCA	ACGCTGACC	GATTTCGGCA
251	AAAAATTTAC	CGACCAAGGT	TTGGACGAAC	CTTCGAAAT	ATGGAATGTC
301	CTCAAAAGCG	AGATGAGCCT	TGTCGGCCCC	GCCGCCCTTT	TGATCGATG

60

5

10

15

```

351 TCTGCGGCTT TACAACAAAT TTCAAAACCG CGGCCACGAA ATGAAACCGG
401 GCATTACCGG CTGGGCGCAG GTCAACGGCGT TCGTGGGACG
451 GAAAAGTTCT CTTGCGATGT TTGGTACACC GACAAATTCA GTTTTGGCT
501 GGATATGAAA ATCCTGTTC TGAAGTCAA AAAAGTCTTG ATTAAGAAAG
551 GCATTTCGGC GCAAGGGGAA GCCACCATGC CCCCTTCGCG GGGGAATCGC
601 AAATTCGCGC TTATCGGCGC GGGCGGACAC GGCAAAGTCG TTGCGGAGCT
651 TGC CGCGCA CTGCGACAT ACGGCGAAAT CGTTTTCTG GACGACCGCA
701 CCCAAGCGCA CGTCAACGGC TTCCCGCTCA TCGGCACGAC GCTGCTGCTT
751 GAAAACAGTT TATCGCCGA ACAATTGCAG ATCACCCTCG CCGTGGGCAA
801 CAACCGCATC CGCGCGCAA TCACCGAAAA CGCGCGCGCG CTCGCGTCAA
851 AACTGCGCGT TCTGATTCT CCGCACGCGA CCGTCTCGCG TTCTGCAATA
901 ATCGGACGAG GCAGCGTCGT AATGGGAGAA GCGCTCGTAC AGGCCGGGAC
951 CGTATTGAAA GACGCGTGA TTGTGAACAC TGCCGCCACC GTCGATCACG
1001 ACTGCCTGCT TGACGCTTTC GtccaCATCA GCCCGGGCGC GCACCTGTG
1051 GGCACACGCG GTATCGGCGA AGAAAGCCGG ATAGCACGCG CGCGCTGCA
1101 CGCGCAGCAG ACAACCGTCG GCAGCGGGGT TACCGCGGCT GACGGGCGCG
1151 TTATCGTATG CGACATCCCG GACGSCATGA CCGTGGCGG CAACCGCGCA
1201 AAGCCCTTA CGGGCAAAA CCCAGAGACC GGGACGGCAT AA

```

This encodes a protein having amino acid sequence <SEQ ID 18>:

20

25

```

1 MSAVKRLFD IASASGLIV LSPVFLVLIY LIRKNLGSVP FTRIRPGKD
51 GKPFKMKVFR SMRDALDSGD IPLEDSERLT DFGKRLRATS LDELPELWNV
101 LKGEMLVGP RPLLMQYLPL YNKFQNRHRE MKPGITGWAQ VNGRNALSWD
151 EKFSQDVWYT DNFSFWLDMK ILFLTVKKVL IKGISAQGE ATMPPFAGNR
201 KLAIVAGAGH GKVAELAA LQTYGEIVFL DRTQGSVNG FVIGITLLLL
251 ENSLSPEQFD ITVAVGNRRI RQITENAAA LGFKLPLVLI PDATVSPSAI
301 IGQGSVVMK AVVQAGSVLK DGVIVNTAAT VDHDCLLDAF VHISPGAHL
351 GNTRIGEESR LGTGACSQQ TTVGSGVTAG AGAIVVDI P DGMTVAGNPA
401 KPIITGKNPRT GTA*

```

This protein shows 86.9% identity in 413 aa overlap with ORF3-1:

30

35

40

45

50

55

60

65

```

10 20 30 40 50 60
orf3-1.pep MSKFFKRLFDIVASASGLIFLSPVFLILYLIIRKNLGSVPVFFQERPGKDGKPFKMKVFR
||| |||||:||||| |||||:||||| |||||:||||| |||||:||||| |||||:|||||
orf3ng MSKAVKRLFDIIASASGLIVLSPVFLVLIYLIIRKNLGSVPVFFTRIRPGKDGKPFKMKVFR
10 20 30 40 50 60
70 80 90 100 110 120
orf3-1.pep SMRDALSDGIPLDGERLTPFGKKLRAASLDELPELWNILKGEMLVGPRLIMQYLPL
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf3ng SMRDALSDGIPLDSERLTDFGKKLRATS LDELPELWNV LKGEMLVGPRLIMQYLPL
70 80 90 100 110 120
130 140 150 160 170 180
orf3-1.pep YDNFQNRHREMKPGITGWAQVNGRNALSWDEKFCADVWYIDHFSLCIDIKILLTVKKVL
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf3ng YNKFQNRHREMKPGITGWAQVNGRNALSWDEKFSQDVWYTDNFSFWLDMKILFLTVKKVL
130 140 150 160 170 180
190 200 210 220 230 240
orf3-1.pep IKGISAQGEATMPPPTGKRKLAVVAGGAGKVVADLAAALGRYRETVFLDRAQGSVNG
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf3ng IKGISAQGEATMPPFAGNRKLAVVAGGAGKVVADLAAALGRYRETVFLDRTQGSVNG
190 200 210 220 230 240
250 260 270 280 290 300
orf3-1.pep FSVIGITLLLENSLSPEQYDVAVVGNRIRRQIAEKAAALGFALPVLIHPDATVSPSAT
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf3ng FSVIGITLLLENSLSPEQFDITVAVGNRIRRQITENAAAALGFALPVLIHPDATVSPSAI
250 260 270 280 290 300
310 320 330 340 350 360
orf3-1.pep VGQGSVVMKAVVQAGSVLKDGVIIVNTAATVDHDCLLNAFVHISPGAHLSGNTHIGESW
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf3ng IGQGSVVMKAVVQAGSVLKDGVIIVNTAATVDHDCLLDAFVHISPGAHLGNTRIGEESR
310 320 330 340 350 360
370 380 390 400 410

```

```

orf3-1.pep      ICTGACSRQQIRIGSRATIGAGAVVVRDVS DGMTVAGNPAKPLPRKNPETSTAX
                |||||:||||| :|||:| |||||:| :| |||||:||||| |||:|:|
orf3ng          ICTGACSRQQTTVGSGVTAGAGAVIVCDIPDGMTVAGNPAKPLTGNPKPTGTAX
                370      380      390      400      410

```

5 In addition, ORF3ng shows significant homology with a hypothetical protein from *B. subtilis*:

```

gnl|PID|e238668 (Z71928) hypothetical protein [Bacillus subtilis]
>gi|1945702|gnl|PID|e313004 (Z94043) hypothetical protein [Bacillus subtilis]
>gi|2635938|gnl|PID|e1186113 (Z99121) similar to capsular polysaccharide
10 biosynthesis [Bacillus subtilis] Length = 202
   Score = 235 bits (594), Expect = 3e-61
   Identities = 114/195 (58%), Positives = 142/195 (72%)

Query: 5 VKRLFDIIASASGLIVLSPVFLVLIYLIRKNLGSVFFIRERPGDKGPKFMVKFSMRD 64
      +KRIFD+ A+ L S + L I ++R +GSPVEF + RPK GKFF + KER+M D
15 Sbjct: 3 LKRLFDLTAALFLLOCTSVIILFTIAVVRLLKIGSPVFFKQVAPGLHGKFFPLTKFRNTMD 62

Query: 65 ALDSGIGLPDSERLTDFGKKLRATSLDELPELWNLKGMESLVGPRPLMLQVLYPLYNKF 124
      DS G LPD RRT G+ +R S+DELF+L NVLKG+SLVVGPRPLLM YLPIY +
20 Sbjct: 63 ERDSKGNLLPDEVRLTKTRGLIRKLKLSIDELPQLLNVLKGDSLVCGRPLMLDYLYLTK 122

Query: 125 QNRHEMKPGITGWAQVNGRNLASWDEKFSQVWYTDNFSFWLDMKILFLTQVKKVLIKEG 184
      Q RRHE+KPGITGWAQ+NGRNA+SW++KF DWY DN+SF+LD+KIL LTV+KVL+ EG
20 Sbjct: 123 QARRHEVKPGITGWAQVNGRNAISWEKKFELDWVYVDNWSFFLDLKLILTVRKVLVSEG 182

25 Query: 185 ISAQGEATMPFFAGN 199
      I T F G+
Sbjct: 183 IQQTNHVTAERFTGS 197

```

The hypothetical product of *yvf* gene shows similarity to EXOY of *R. meliloti*, an
 30 exopolysaccharide production protein. Based on this and on the two predicted transmembrane
 regions in the homologous *N. gonorrhoeae*, it is predicted that these proteins, or their
 epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 4

The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 19>:

```

35 1 ..AACCATATGG CGATTTCAT CGACGAATAC GCGGCGACAT CCGGCTTGCT
   51 CACCTTTGAA GACATCATCG ACRAATCGT CGCGGAATC GAAGACGACT
   101 TTGACGAAGA CGATACGCCG GACAAATATCC ATGCCCTTTC TTGACACAGC
   151 TGGCGCATCC ATGCACCTAC GAAATCGAA GACATCAACA CCTTCTTCGG
40 201 CACGGAATAC AGCATCGAAG AAGCCGACAC CATT.GCGGG CCGTGTCATT
   251 CAAGAGTTGG GACATCTGCC CGTGCGCGCG GAAAAAGTCC TTATCGGCGG
   301 TTTGCAGTTC ACCGTCGCAC GCGCGCGCAA CCGCGCGCTG CATACGCTGA
   351 TGGCGACCCG CGTGAAGTAA GC..... ..ACCGC GCTTTCTGCA
   401 CAGTTTAG

```

This corresponds to amino acid sequence <SEQ ID 20; ORF5>:

```

45 1 ..NHMAIVIDEY GGTSGLVTFE DIIEQIVGEI EDEFDEDDSA DNIHAVSSDT
   51 WRIHAATEIE DINTFFGTEY SIEEADTIXR PGHSRVGTSA RARRKSPYRR
   101 FAVHRRTRRQ PPPAYADGDP REVS....XR RFCTV*

```

Further sequence analysis revealed the complete DNA sequence to be <SEQ ID 21>:

```

50 1 ATGAGCGGCG CACAACCGAA AACGAATTTT TTSTAAGCGC TGATTGCGCG
   51 ACTGCGCCGC GACCCGAGT CCGCCGAGT CGTATTATAC CTGCTTCGCG
   101 AGCGGACGCA CGAGGACAT TTTCATCGCG ATAGCTGCTT AGATTGTGAA
   151 AAGATCTCTG ATTTTTCGGA TTGGAAGTGC CGGCACGCGA TGATTACGCG
   201 CAGCCGATAT AACGCTTTAA AAGAAAACGA CAGCATCGAG CCGATCACCG
   251 CCTACGTTAT CGATACCGCC CATTGCGCCT TCCCGTCAT CGCGGAAGAC

```

5 301 AAGACGAAG TTTTGGCGAT TTTGCAGGCC AAGACCTGC TCAATATAT
 351 GTTAAACCC GAGCAGTTCC ACCTCAAATC CATTTCTCCG CCGCGCGTCT
 401 TCGTCCCGGA AGCRAATCG CTGACCGCCC TTTTAAAGA GTTCCGGAA
 451 CAGCGCAACC ATATGGCGAT TGTATCGAGC GAATACGGCG GCACATCCGG
 501 CTTGGTCACC TTTGAAGACA TCATCGAGCA AATCGTCGGC GAAATCGAAG
 551 ACGAGTTTGA CGAAGACGAT AGCGCCGACA ATATCCATGC CGTTTCTTCC
 601 GAACGCTGGC GCATCCATGC AGCTACCGAA ATCGAAGACA TCAACACCTT
 651 CTTGGCGCAG GAATACAGCA GCGAAGAAGC CGACACCATT CGGCTTGCTC
 10 701 ATTCAGAGT TGGGACATCT GCCCGTGGCG GCGGAAAAGG TCCTTATCGG
 751 CGGTTTGCG TTCACCGTCG CACGCGCCGA CACCGCCGCG CTGCATACGC
 801 TGATGGCGAC CCGCGTGAAG TAAGCACCGC CGTTTCTGCA CAGTTTAGGA
 851 TGACGCTACG GCGTTTCTCT GTTTCATCC GCCCATCCG CCAACATAA

This corresponds to amino acid sequence <SEQ ID 22; ORF5-1>:

15 1 MDGAQPKTNF FERLIARLAR EPDSAEDVLN LLRQAHEQEV FDADTLRLLE
 51 KVLDFSDLEV RDMITRSRM NVLKENDSIE RITAYVIDTA HSRFFVIGED
 101 KDEVILGILHA KDLLKYMFPN EQFHLSILR PAVFVPEGKS LTALLKEFRE
 151 QRNHMAIVID EYGGTSLGLV FEDIIEQIVG DIEDFEDEDD SADNIHAVSS
 201 ERWRHIAATE IEDINAFFTG EYSSEADTI RPHGSRVGTG ARARRKSPYR
 251 RFAVHRRTRR QPPPAYADG PREVSTAVSA QFRMTVRAFV VSIRPIRQT*

20 Further work identified the corresponding gene in strain A of *N.meningitidis* <SEQ ID 23 >:

25 1 ATGGACGGCG CACAACCGAA AACAAATTTT TTNNAACGCC TGATTGCCCG
 51 ACTCGCCCGC GAACCCGATT CCGCGGAAGA CGATTAGACC CTGTTGGCGC
 101 AAGCGCACGA ACAGGAAGTA TTTGATCGGG ATACCGTTT AAGATTGGAA
 151 AAGATCCTCG ATTTTCTGA TTTGGAAGT CGCGACGGGA TGATTACGG
 201 CAGCGGTATG AACGTTTAA AAGAAACGGA CAGCATCGAA CGCATCACCG
 251 CCTACGTTAT CGATACCGCC CATTCGCGCT TCCTCGCAT CGGTGAAGAC
 301 AAGACGAAG TTTTGGGTAT TTTGCAGGCC AAGACCTGC TCAATATAT
 351 GTTCAACCCC GAGCAGTTCC ACCTCAAATC GATATTGCGC CTCGCGCTCT
 401 TCGTCCCGGA AGGCAATCG CTGACCGCCC TTTTAAAGA GTTCCGGGAA
 451 CAGCGCAACC ATATGGCAAT CGTCATCGAC GAATACGGCG GCACGCTGGG
 501 TTTGGTAACT TTTGAAGACA TCATCGAGCA AATCGTCGCG GACATCGAAG
 551 ATGAGTTTGA CGAAGACGCA AGCGCGGACA ACATCCAGCG CGTTCCGCGC
 601 GAACGCTGGC GCATACAGCG GGTACCGATA ATCGAAGACA TCAACGCTT
 651 TTTCCGACG GAATACAGCA GCGAAGAAGC CGACACCATC GGCGGCGNTG
 701 TGCATTGAGC AATTGNACA CTTGCCGCTG CGCGCGGAAA AAGTCMTTAT
 751 CGCGGNNTTG CANTTCACNG TCGCCNGCGC NGACAAACCG CGCCTGCATA
 801 CGCTGATGGC GACCCGCGTG AAGTAGCTC CGCCGTTTCT GTACAGTTTA
 851 GGATGACGGT ACGGGCGTTC TCTGTTTCAA TCGCGCCCAT CCGGCANACA
 901 TAA

40 This encodes a protein having amino acid sequence <SEQ ID 24; ORF5a>:

45 1 MDGAQPKTNF XRRLIARLAR EPDSAEDVLT LLRQAHEQEV FDADTLRLLE
 51 KVLDFSDLEV RDMITRSRM NVLKENDSIE RITAYVIDTA HSRFFVIGED
 101 KDEVILGILHA KDLLKYMFPN EQFHLSILR PAVFVPEGKS LTALLKEFRE
 151 QRNHMAIVID EYGGTSLGLV FEDIIEQIVG DIEDFEDEDD SADNIHAVSA
 201 ERWRHIAATE IEDINAFFTG EYSSEADTI RPHGSRVGTG PARARRKSYX
 251 RRXAXHXRKR QPPPAYADG DPREVSSAVS VQFRMTVRAFV SVSIRPIRXT
 301 *

The originally-identified partial strain B sequence (ORF5) shows 54.7% identity over a 124aa overlap with ORF5a:

50 orf5.pep 10 20 30
 NHMAIVIDEYGGTSLGLVTFEDIIEQIVGEI
 orf5a FHLKSLRPVAVFVPEGKSLTALLKEFREQRNHMAIVIDEYGGTSLGLVTFEDIIEQIVGDI
 130 140 150 160 170 180
 55 orf5.pep 40 50 60 70 80 90
 EDEFEDEDSADNIHAVSSDTRWIRHIAATEIEDINTFFGTEYSIEEADTIKRPGHRSRVGTSA
 orf5a EDEFEDESADNIHAVSAERWIRHIAATEIEDINAFFTGTEYSSEADTIIGKGXHSIGITPA
 190 200 210 220 230 240
 60

		100	110	120	130	
orf5.pep		RARRKSPYRRFAVHRRTTRQPPPAYADGDPREVSSXXXXRRFCTV				
5	orf5a					
		RARRKSKYRRKAXHXRKXQPPPAYADGDPREVSSAVSQFRMTVRAFSVSIRPIRXTX				
		250	260	270	280	290 300

The complete strain B sequence (ORF5-1) and ORF5a show 92.7% identity in 300 aa overlap:

		10	20	30	40	50	60
10	orf5a.pep	MDGAQPKTNFXXRLIARLAREPDSAEVDLTLLRQAHEQEVFDADTLRLLEKVLDFSDLEV					
	orf5-1	MDGAQPKTNFXXRLIARLAREPDSAEVDLTLLRQAHEQEVFDADTLRLLEKVLDFSDLEV					
		10	20	30	40	50	60
		70	80	90	100	110	120
15	orf5a.pep	RDAMITSRMNVLKENDSIERITAYVIDTAHSRFPVIGEDKDEVLSLHAKDLKYMFP					
	orf5-1	RDAMITSRMNVLKENDSIERITAYVIDTAHSRFPVIGEDKDEVLSLHAKDLKYMFP					
		70	80	90	100	110	120
		130	140	150	160	170	180
20	orf5a.pep	EQFHLKSLIRPAVFVPEGKSLTALLKEFREQNHMAIVIDEYGGTSGLVTFEDIEIQIVG					
	orf5-1	EQFHLKSLIRPAVFVPEGKSLTALLKEFREQNHMAIVIDEYGGTSGLVTFEDIEIQIVG					
		130	140	150	160	170	180
		190	200	210	220	230	240
25	orf5a.pep	DIEDEFDEDESADNIHAVSAERWRIHAATEIEDINAFGGTEYSSEADTIGGXGHSIGT					
	orf5-1	EIEDEFDEDDADNIHAVSSERWRIHAATEIEDINAFGGTEYSSEADTIRF-GHSRVGT					
		190	200	210	220	230	
30		250	260	270	280	290	300
	orf5a.pep	PARARRKSKYRRKAXHXRKXQPPPAYADGDPREVSSAVSQFRMTVRAFSVSIRPIRXT					
	orf5-1	SARARRKSPYRRFAVHRRTTRQPPPAYADGDPREVSTAVSAQFRMTVRAFSVSIRPIRQT					
		240	250	260	270	280	290

Further work identified the a partial DNA sequence in *N.gonorrhoeae* <SEQ ID 25> which encodes

a protein having amino acid sequence <SEQ ID 26; ORF5ng>:

40	1	MDGAQPKTNF	FERLIARLAR	EPDSAEVDLN	LLRQAHEQEV	FDADTLRL
	51	KVLDFAELV	RDAMITSRM	NVLKENDSIE	RITAYVIDTA	HSRFPVIGED
	101	KDEVGLILHA	KDLLKYMFP	EQFHLKSVLR	PAVFVPEGKS	LTALLKEFRE
	151	QRNHMAIVID	EYGGTSGLV	FEDIEIQIVG	DIEDEFDEDE	SADDIHSVSA
	201	ERWRIHAATE	IEDINAFGT	EYGGSEADTI	RRLGHSIGT	PARARRKSEY
	251	RRFAVHRPR	RQFFFAHADG	DPREVSRACP	HRRFCT*	

45 Further analysis revealed the complete gonococcal nucleotide sequence <SEQ ID 27> to be:

	1	ATGGAACGGCG	CACAACCGAA	AACAAATTTT	TTTGAACGCG	TGATTGCCCG
	51	ACTCGCCGCG	GAACCCGATT	CCGCCGAAGA	CSTATTAAAC	CTGCTCGCG
	101	AGGCGCAGCA	ACAGGAAGTT	TTTGATGCGC	ACACACTGAC	CCGCTCGGAA
	151	AAAGTATTGG	ACTTTGCGCA	GCTGGAASTG	CCGATATGCG	TGATTACCGC
50	201	CRGCGCATC	ACATATTGCA	ARGAAACGA	CAGCATCGAA	CGCATACCGC
	251	CCTACGTCAT	CGATACCGCC	CATTGCGGCT	TCCCGTCTAT	CGSGCGAAGAC
	301	AAAGACAGAG	TTTGGGCGAT	TTTGACCGCC	AAAGACCTGC	TCAATATAT
	351	GTTCACCCCG	GAGCAGTTCC	ACCTGAAATC	CGTCTTGGCG	CCTGGCGTTT
	401	TCGTGCCCCG	AGGCAAATCT	TTGACCGCCC	TTTAAAGAAG	GTTCGCGAAG
55	451	CAGCGCAACC	ATATGGCAAT	CGTCATCGAC	GAATACGGCG	GCACGTCCGG
	501	TTTGGTCAAC	TTTGAAGACA	TCAATCGAGA	AATCGTCGGT	GACATCGAAG
	551	ACGAGTTTGA	CGAAGACGAA	AGCGCGcagc	acatCCATCT	cgTTTccgCC
	601	GAACGCTGGC	GCATCCcagc	ggctaCCGAA	ATCAGAGaca	TCAACGCCIT
	651	TTTCGTTAGC	Gaatcacgca	gcgaagaagc	cgacaccatc	cggcggtTG
	701	GTCATTTCAG	AATTGGGACA	CTGCCCCGTG	CGCGCGGAAA	AAGTCTTAT
60	751	cggcggtTTG	Cagttcaacy	tCGCCCGCGC	CGACAACCGC	CGCTCGCACA
	801	CGCTGATGGC	GACCCGCGTG	AAGTAAGCAG	AGCCTGCGCG	AcccggtT

851 CTGCacAGTT TAGGatgACG gtaCGTGTGTTTCTGTTC AATCCGCCCC
901 ATCCGCCAAA CATAA

This encodes a protein having amino acid sequence <SEQ ID 28; ORF5ng-1>:

5 1 MDGAQPKTNF FERLIARLAR EPDSAEDVIN LLRQAHEQEV FDADTLTRLE
51 KVLDFAELEV RDMITRSRM NVLKENDSIE RITAYVIDTA HSRFPVIGED
101 KDEVILGILHA KDLLKYMFPN EQFHLSVLR PAVFVPEGKS LTALLKEFRE
151 QRNHMAIVID EYGGTSGLVT FEDIIEQIVG DIEDEFDEDE SADDIHSVSA
201 ERWRIHAATE IEDINAFPGT EYGSEEDATI RRLGHSIGGT PARARRKSPY
251 RRFVHRRPR RQPPPAHADG DPREVSRACP TAVSAQFRMT VRSFVSIRP
10 301 IRQT*

The originally-identified partial strain B sequence (ORF5) shows 83.1% identity over a 135aa overlap with the partial gonococcal sequence (ORF5ng):

15 orf5 NHMAIVIDEYGGTSGLVTFEDIIEQIVGEI 30
orf5ng FHLKSVLRPAVFVPEGKSLTALLKEFREQRNHMAIVIDEYGGTSGLVTFEDIIEQIVGDI 182
orf5 EDEFDEDDSDADNIHAVSSDTRIHAATEIEDINTFFGTEYSIEEADTIRLGHSGHVRGTS 90
20 orf5ng EDEFDEDESADDIHSVSAERWRIHAATEIEDINAFGT EYGSEEDATIRRLGHSIGGTGA 242
orf5 RARRKSPYRRFAVHRRTRRQPPPAYADGDPREVSRP---RRFCTV 131
orf5ng RARRKSPYRRFAVHRRTRRQPPPAHADGDPREVSRACPHRRFCTV 287

The complete strain B and gonococcal sequences (ORF5-1 & ORF5ng-1) show 92.4% identity in

304 aa overlap:

orf5ng-1.pep 10 20 30 40 50 60
MDGAQPKTNF FERLIARLAR EPDSAEDVIN LLRQAHEQEV FDADTLTRLEKVLDFAELEV
orf5-1 MDGAQPKTNF FERLIARLAR EPDSAEDVIN LLRQAHEQEV FDADTLTRLEKVLDFAELEV
30 10 20 30 40 50 60
orf5ng-1.pep 70 80 90 100 110 120
RDMITRSRM NVLKENDSIE RITAYVIDTA HSRFPVIGED KDEVILGILHAKDLLKYMFPN
orf5-1 RDMITRSRM NVLKENDSIE RITAYVIDTA HSRFPVIGED KDEVILGILHAKDLLKYMFPN
35 70 80 90 100 110 120
orf5ng-1.pep 130 140 150 160 170 180
EQFHLSVLR PAVFVPEGKSLTALLKEFREQRNHMAIVIDEYGGTSGLVT FEDIIEQIVG
orf5-1 EQFHLSVLR PAVFVPEGKSLTALLKEFREQRNHMAIVIDEYGGTSGLVT FEDIIEQIVG
40 130 140 150 160 170 180
orf5ng-1.pep 190 200 210 220 230 240
DIEDEFDEDESADDIHSVSAERWRIHAATEIEDINAFGT EYGSEEDATIRRLGHSIGGT
orf5-1 DIEDEFDEDDSDADNIHAVSSERWRIHAATEIEDINTFFGTEYSIEEADTIRP-GHSRVGT
45 190 200 210 220 230
orf5ng-1.pep 250 260 270 280 290 300
PARARRKSPYRRFAVHRRTRRQPPPAHADGDPREVSRACPTAVSAQFRMTVRSFVSIRP
orf5-1 SARARRKSPYRRFAVHRRTRRQPPPAYADGDPREVSR---TAVSAQFRMTVRSFVSIRP
50 240 250 260 270 280 290
orf5ng-1.pep IRQTX
55 11111
orf5-1 IRQTX
60 300

Sbjct: 70 RD MI RS+M LK N +++ +I++AHSRFPVI EDKD + GIL AKDLL +M +
 RDIMIPRSQMITLKRNTLDECLDVIIESAHSRFPVISEDKDHIEGILMAKDLLPFMRSD 129

Query: 120 PEQPHLKSILRPVAVFPEGKSLTALLKEFREQRNHMAIVIDEYGGTSGLVTFEDIEIQIV 179
 E F + +LR AV VPE K + +LKEFR QR HMAIVIDE+GG SGLVT EDI+E IV

Sbjct: 130 AEAFSMDKVLRAVAVVPEKRVDRMLKEFRSQRYHMAIVIDEFGVSGLVITIEDILELIV 189

Query: 180 GDIEDEFDEDESADNIHVAESAERWRIHAATEIEDINAFPGTSEYSEADT 229
 G+IEDE+DE++ D +S W + A IED N FGT +S EE DT

Sbjct: 190 GEIEDEYDEEDDID-FRQLSRHTWTVRALASIEDNEAFGTHFSDEEVD 238

Based on this analysis, including the amino acid homology to the TlyC hemolysin-homologue from *H. influenzae* (hemolysins are secreted proteins), it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae* are secreted and could thus be useful antigens for vaccines or diagnostics.

ORF5-1 (30.7kDa) was cloned in the pGex vector and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 2A shows the results of affinity purification of the GST-fusion protein. Purified GST-fusion protein was used to immunise mice, whose sera were used for Western blot analysis (Figure 1B). These experiments confirm that ORF5-1 is a surface-exposed protein, and that it is a useful immunogen.

Example 5

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 29>:

```

1 ATGCGCGCGC GCAGGCGCGA TTCGTTACC GTGCAGATTA TCGAAGGTTT
51 GCGTTTTTCG CATATGAGGA AGTTCATCGA CGCAACGCC CACATCGGAC
101 ACGACACCAA AGGCTGGAGC AATGAAAAAC TGATGGCGGA AGTTGCGCCC
151 GATGCCTTCA GCGGCAATCC TGAAGGGCAG TTTTCCCCG ACAGCTACGA
201 AATCGATCGG GCGGCGAGTG ATTTGCAGAT TTACCAAAAC GCCTACAAGG
251 GCGATGCACG GCCGCTGAA TGAAGGCATG GGAAGCAGG CAGGACGGGC
301 TGCCTTATAA AAACCTTAT GAAATGCTGA TTATGGCGA CCGTGTGCAA
351 AAGGAAACAG GGCATGAAGC CGAAGCAGC CATGTGCTT CCGTCTCGT
401 CACCGCGCTG AAAATCGGTA TGCGCTGCA AACCGAAGC TCGCTGATT
451 ACGGCATCGG TCGGCGATAC AAGGCGAACA TCCGTAAAC CGACCTGCGC
501 CCGGACAGC CGTACACACG CTACACGCGC GCGGCTCTG CGCCACCCC
551 GATTGCGCTG CCC..

```

This corresponds to the amino acid sequence <SEQ ID 30; ORF7>:

```

1 MRGGRPDVST VQIIIEGRFS HMRKVIDATP DIGHDTKGWS NEKLMAEVAP
51 DAFSGNPEQG FFPDSYEIDA GGSDLQIYQT AYKAMQRLRN EAWESRQDGL
101 PYKNPYEMLI MAXLVEKETG HEAXXDHVAS VEVNRLKIGM RLQTXSVIY
151 GMGAAYKGKI RKADLRDPT YNTYTRGLP PTPIALP..

```

Further sequence analysis revealed the complete DNA sequence <SEQ ID 31>:

```

1 ATGTTGAGAA AATTGTTGAA ATGCTCTGCC GTTTTTTGA CCGTGTGCGC
51 AGCCCGTTTT CGCGCGCTGC TTTTGTGTTT TAAGGATAAC GGCAGGGCAT
101 ACCGAATCAA AATTGCGAAA AACAGGGTGA TTTCGTCGGT CGCGAGGAAA
151 CTTGCGCAAG ACCGCATCGT GTTCAGCAGG CATGTTTTGA CGCGCGCGGC
201 CTACGTTTTG GGTGTGCACA ACAGGCTGCA TACGGGAGC TACAGATTGC
251 CTTCGGAAGT GTCTGCTTGG GATATCTTGC AGAAAATCGG CGCGCGCAGG
301 CCGGATTCGG TTACCGTGGG GATATTCGAA GTTTCGCTT TTTTGCACA
351 GAGGAAATGC ATCAGCGCAA CCGCCGACAT CGGACACGAC ACCAAGCGGT
401 GGAGCAATGA AAAACTGATG GCGGAAGTTG CCGCGCATGC CTTACGCGCG
451 AATCCTGAAG GCGAGTTTTT CCGGACAGC TACGAATATG ATGCGGCGCG

```

501	CAGTGAATTGG	CAGATATTACC	AAGCCGCGCTA	CAGGCGGCTA	CAACGCGCGC
551	TCGAGGAAGC	ATCGGAAAGC	AGGCAGCAGC	CGCTGCGTTA	TAAAGAACCT
601	TATGAAGCT	TGATTATTGCG	GAGCCGTGTG	GAAAGGAAGA	CAGGCGATGA
651	AGCCACGACCG	GACCATGTGCG	CTTCGCGTCT	CGTCAACCGC	CTGAAGATAT
701	TCTTCGCGCT	CAACACCGAC	CGGTCCGTGA	TTTACGCGAT	GGGTGCGGCA
751	TATAGAGGCA	AAATCGTATA	AGCGCAGCTG	CGCCGCGACA	CGCGCTACAA
801	CACCTACACG	CGAGTGGCGTC	CGCCGCGCAAC	CCCGATTGCG	CTGCCCGGCA
851	AGCGCGCACT	CGGTCGCGCC	TGCCATCGGT	CGCGCAAAA	ATACCTGTAT
901	TTTGCTGTCA	AATGAAGGAG	CACGGCGTTG	AGGCAGTTCA	CGCATGTATT
951	GCAGCGAAGC	AATGCGCGCG	TCGCGABATA	TATTTTGAAG	ABATTA

This corresponds to the amino acid sequence <SEO ID 32: ORF7-1>:

1	MLRKLLKWSA	VFLTVSAAVF	AALLFVPEKN	GRAYRIKIAK	NQGISVGRK
51	LAEDRVTFSR	HLTAAYYLV	GDVNRLHGTG	YRLPSEVSAN	DLQKAMPGRG
101	PDSTVQIIE	SVRSSHMVK	VDNTPDIGH	TKGSENEKIM	AEVAPDASGG
151	YMEQGFPPDS	YIEDAGGSLD	QHYQATYAKM	QRRLEAGRES	RQDGLPYKNP
201	NEPLIMASLV	EKETGHEADR	DIIVATVFNRL	LIGMRLOQTD	PSVIYGMGAA
251	YKGKIRKADL	RRDTPNNYTG	RGGLPTPTIA	LPGKAALDAA	AHFSGEKYLY
301	FVSKMDGTGL	SQSFOLDTEH	NAARVYKILK	K*	

Computer analysis of this amino acid sequence gave the following results:

20 Homology with hypothetical protein encoded by *yceg* gene (accession P44270) of *H.influenzae*
ORF7 and yceg proteins show 44% aa identity in 192 aa overlap:

ORF7	1	MGRGGPSDVTVQITIEGRSFRSHMKRVIDATPDIGHDTKGWSNEKLA-----EVAPDAFSG	55
ycg	102	LNSGKEVQVFNVWKEIGKTFDKWRKDLNAPHVLQTLKDKSNEEIFALLDLPDIGNLEK	161
ORF7	56	NPEGQFSDPSYSYIEDAGSGDLQTYQYAKMORRINEAWESRQDGLPYKNPYEMIIMAXLV	115
ycg	162	NVEGWLVPDTYNTYTPKSTDLLELKKRSAEKMKALNKAWNERDEDDKIPADLPYEMIILASIV	221
ORF7	116	EKETGEAAXXDHVASVFNRLKIGMRLLQTKXSXSVIYMGAAAYKGKIRKDLRDTPYNTYT	175
ycg	222	EKETGIANERAKVASVFINRLKAKMKLQDTPVTVYMGWENYGNIRKKDELTKTPTNTYTV	281
ORF7	176	RGGLPPTTPIAM	187
ycg	282	IGLPTTPIAM	293

The complete length YCEG protein has sequence:

40

1	KKFKLPIAILL	LILLILAGVAS	FSYYKMTFV	KTPVNVQADE	LLTIERGTTS
51	LKTLRFEQF	KLADGKLLP	YLKLKLPEN	PHKQVTLDS	NVKTVDQLDL
101	LLNSGKVEFQ	NWMLBKGTF	KWRKRLDPA	PIKVQTLKDK	SNEEIFALDL
151	LPDIGNQLL	KNVGUEWLYP	DVTYTPKST	LELLKRSAR	MKKNKAWN
201	EREDDLPIAN	PYEMILLIL	VEKTEGIANE	RKAVSVFIN	RLKAKMYKLT
251	DPVTIYGMGE	NYVNGIRKKD	LETKTPEYNT	VIDGLPPTFI	AMPSSSLQA
301	VANPEKTDFF	FFYFADSGSG	KPTRNLNEHD	KAVOEYLRWY	RSOKNA

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF7 shows 95.2% identity over a 187aa overlap with an ORF (ORF7a) from strain A of *N. meningitidis*:

50 orf7.pep 10 20 30
MRGGRPDSVTVTQIEGSRFSSHMRKVIDATP
orf7a AAYVLGVHNRLLHTGYRLPEVSAWDILKMRGGRPDSVTVTQIEGSRFSSHMRKVIDATP
70 80 90 100 110 120

55 orf7.pep 40 50 60 70 80 90
DIGHDTKGWSNEKLMAEVAPDASFGNPEGQFPDSEYIDAGGSDLQIYYQAYKAMQRRLN

orf7a	DIEHDTKGWSNEKLMAEVAPDAFSGNPEGQFFPDSEYIDAGGSDRLRIYQIAYKAMQRRRLN	130	140	150	160	170	180
5	orf7.pep	100	110	120	130	140	150
orf7a	EAWESRQDGLPYKNPYEMLIMAKLVEKETGHEAXXDHVASVFVNRNLKIGMRLQTXKSVIY 	160	170	180	190	200	210
orf7a	EAWESRQDGLPYKNPYEMLIMASLIEKETGHEADRHVASVFVNRNLKIGMRLQTDPSVIY 	220	230	240	250	260	270
10	orf7.pep	160	170	180	190	200	210
orf7a	GMGAAYKGKIRKADLRRTDPYNTYTRGGLPPTPIALP 	220	230	240	250	260	270
15	orf7a	250	260	270	280	290	300
orf7a	DGTGLSQFSDHLTEHNAAVRKYLKXX 310 320 330						

The complete length ORF7a nucleotide sequence <SEQ ID 33> is:

20	1	ATGTTGAGAA	AATGTTGAA	ATGCTCGCC	GTTTTTTGA	CGTATCGGC
	51	AGCGGTTTC	GCGCGCTGC	TTTTGCTCC	TAAAGACAC	GGCAGGCGAT
	101	ACAGGATTAA	AATTGCGAAA	AACGAGGTA	TTTCGTCGT	CGCGAGGAAA
	151	CTTGCCGAAG	ACCGCATCGT	GTTACGACG	CATGTTTTGA	CGCGCGCGCG
	201	CTACGTTTTG	GGTGTGCACA	ACAGGCTGCA	TACGGGACG	TACAGACTCG
	251	CTTCGGAAGT	GTCGTCTGG	GATATCTTG	AGAAAATGCG	CGCGCGCAGG
25	301	CGGATTCGG	TTACCGTGCA	GATTATCGAA	GGTTCGCGT	TTTCGCATAT
	351	GAGGAAAGTC	ATCGACGCAA	CGCCGACAT	CGAACACGAC	ACCAAAGGCT
	401	GGAGCAATGA	AAAACGTGAT	CGCGAAGTTG	CCCTGATGC	CTTCAGCGCG
	451	AATCCTGAAG	GGCAGTTTTT	CCCGACACG	TACGAAATCG	ATCGCGCGCG
	501	CAGCGATTTA	CGGATTTACC	AAATCGCCTA	CAAGCGCATG	CAACCGCCAG
30	551	TGATGAGGC	ATGGGAAAGC	AGGCAGGACG	GGCTGCCTTA	TAAAAACCTT
	601	TATCAAATGC	TGATTTATGG	GAGCCTGATC	GAAAAGGAAA	CAGGCGCATGA
	651	AGCGACGCG	CAGCATCTCG	CTTCCGCTT	CGTCAACGCG	CTGAAATATG
	701	GTATGCGCCT	GCAAACCCGAC	CCGTCGCTGA	TTTACGGCAT	GGTGCGCGCA
35	751	TACAAGGCGA	AAATCCGTAA	AGCCGACCTG	CGCCGCGACA	CGCCGTACAA
	801	CACCTACACG	CGCGGCGGTC	TGCCGCCAAC	CGCATCGCG	CTGCCCGGCA
	851	AGGCGGCACT	CGATGCGCGC	GCCCATCCGT	CCGGTGAAA	ATACCTGTAT
	901	TTCTGTGCCA	AAATGGACGG	TACGGGCTTG	AGCCAGTTCA	GCCATGATTT
	951	GACCGAACAC	AACGCGCGCG	TTCCGAAATA	TATTTTGAAA	AAATAA

This is predicted to encode a protein having amino acid sequence <SEQ ID 34>:

40	1	MLRKLKWSA	VFLTVSAAVF	AALLFVPKDN	GRAYRIKIAK	NQGISSVGRK
	51	LAEDRIVFSR	HVLTAAYVL	GVNRLHTGT	YRLPSEVS	AWDLQKMRGR
	101	PDSVTVQIIE	GSFRFSHMRK	IDATPDIEHD	TKGWSNEKLM	AEVAPDAFSG
	151	NPEGQFFPDS	YEIDAGGSDL	RLYQIAYFAM	QRRLINEAES	RQDGLPYKNP
45	201	YEMLMASLI	EKETGHEADR	HDVASVFVNR	NLKIGMRLQTD	PSVIYGMGAA
	251	YKGIKIRKADL	RRDTPYNTYT	RGGLPPTPIA	LPGKAALDAA	AHPSGGEKYL
	301	FVSKMDGTGL	SQFSDHLTEH	NAAVRKYLK	K*	

A leader peptide is underlined.

ORF7a and ORF7-1 show 98.8% identity in 331 aa overlap:

50	orf7a.pep	10	20	30	40	50	60
	orf7-1	10	20	30	40	50	60
55	orf7a.pep	70	80	90	100	110	120
	orf7-1	70	80	90	100	110	120
60		130	140	150	160	170	180

	orf7a.pep	IDATPDIEHDTKGWSNEKLMAEVAPDAFSGNPEGQFFPDSYEIDAGGSDLRIYQIAYKAM	
	orf7-1	IDATPDIGHDTKGWSNEKLMAEVAPDAFSGNPEGQFFPDSYEIDAGGSDLQIYQTAYKAM	
5		130 140 150 160 170 180	
	orf7a.pep	190 200 210 220 230 240	
	orf7-1	190 200 210 220 230 240	
10		250 260 270 280 290 300	
	orf7a.pep	PSVIYGMGAAYKGKIRKADLRDTPYNTYTRGGGLPTPIALPGKAALDAAAHPSGEKYL	
15	orf7-1	PSVIYGMGAAYKGKIRKADLRDTPYNTYTRGGGLPTPIALPGKAALDAAAHPSGEKYL	
		250 260 270 280 290 300	
20	orf7a.pep	310 320 330	
	orf7-1	FVSKMDGTGLSQFSHDLTEHNAAVRKYLK	
		310 320 330	

Homology with a predicted ORF from *N.gonorrhoeae*

25 ORF7 shows 94.7% identity over a 187aa overlap with a predicted ORF (ORF7.ng) from *N. gonorrhoeae*:

	orf7	MRGGRPDSVTVQIIEGSRFSHMRKVIDATPDIGHDTKGWSNEKLMAEVAPDAFSGNPEGQ	60
	orf7ng	MRGGRPDSVTVQIIEGSRFSHMRKVIDATPDIGHDTKGWSNEKLMAEVAPDAFSGNPEGQ	60
30	orf7	FFPDSYEIDAGGSDLQIYQTAYKAMQRRRNEAWESRDGLPYKNPYEMLIMAXLVEKETG	120
	orf7ng	FFPDSYEIDAGGSDLQIYQTAYKAMQRRRNEAWESRDGLPYKNPYEMLIMASLIEKETG	120
35	orf7	HEAXXDHVASVFNRLKIGMRLQTXSVIYGMGAAYKGKIRKADLRDTPYNTYTRGGGLP	180
	orf7ng	HEADRDHVASVFNRLKIGMRLQTDPSVIYGMGAAYKGKIRKADLRDTPYNTYTRGGGLP	180
40	orf7	ETPIALP	187
	orf7ng	ETRIALPGKAAMDAAHPSGEKYL FVSKMDGTGLSQFSHDLTEHNAAVRKYLK	236

An ORF7ng nucleotide sequence <SEQ ID 35> is predicted to encode a protein having amino acid sequence <SEQ ID 36>:

45	1	MRGGRPDSVT	VQIIEGSRFS	HMRKVIDATP	DIGHDTKGWS	NEKLMAEVAP
	51	DAFSGNPEGQ	FFPDSYEIDA	GGSDLQIYQT	AYKAMQRRRLN	EAWAGRQDGL
	101	PYKNPYEMLI	MASLIEKETG	HEADRDHVAS	VFNRLKIGM	RLQTDPSVIY
	151	GMGAAYKGKI	RKADLRDTP	YNTYTRGGGLP	ETRIALPGKA	AMDAAHPSG
	201	EKYL FVSKM	DGTGLSQFSH	DLTEHNAAVR	KYLK	*

Further sequence analysis revealed a partial DNA sequence of ORF7ng <SEQ ID 37>:

50	1	..taccgaatca	AGATTGCCAA	AARTCAGGGT	ATTTCTGTCGG	TCCGCAGGAA
	51	ACTTGCcgaA	GACCGCATCG	TGTTACAGCAG	GCATGTTT	ACAGCGCGCG
	101	CCTACGTTTT	GGGTGTGCAC	AACAGGCTGC	ATACGGGGAC	gTACAGATTG
	151	CCTTCGGAA	TGTCTGCTTG	GGATATCTTG	CAGAAAATGC	GCGGCGCGCAG
55	201	GCGGATTC	GTACCGTGC	AGATTATCGA	AGGTTCCGCGT	TTTTCGCAT
	251	TGAGGAANAGT	CATCGACGCA	ACGCCGACG	TCGGACACGA	CACCAAGGC
	301	TGGAGCAATG	AAAAACTGAT	GGCGAGCTT	GCGCGCATG	CTTTCAGCG
	351	CAATCCTGAA	GGGCAGTTTT	TTCCCGACAG	CTACGAATC	GATGCGGGCG
	401	GCAGCGATT	GCAGATTTC	CAAACCGCCT	ACAAGCGCAT	GCAAGCCGCG
	451	CTGAACGAGG	CATGGGCAGG	CAGGCAGCAG	GGGCTGCCTT	ATAAAAACCC
60	501	TTATGAAATG	CTGATTATGG	CGAGCCTGAT	CGAAAAGGAA	ACGGGCGCATG

551 AGGCCGACCG CGACCATGTC GCTCCGTCT TCGTCAACCG COTGAAATTC
 601 GGTATGCGCC TGCAAAACCGA CCGTCCCGT ATTTACGGCA TGGGTGCGGC
 651 ATCAAGGGC AAAATCCGTA AAGCCGACCT GCGCCGGAC ACGCCGTACA
 701 aCacCTAtac gggcgggggc ttgcccgcaa cccggattgc gctgccggcC
 751 Aagcgcgcaa tggatgcgcg cgccaccccg tccggcgaaAaatactcgtTa
 801 ttctgtctcC AAAATGGACG GCACGGGCTT GAGCCAGTTC AGCCATGATT
 851 TGACCGAACA CAACGCGCCg gTcCGCAAAAT ATATTTTGAA AAAATAA

This corresponds to the amino acid sequence <SEQ ID 38; ORF7ng-1>:

1 ..YRIKIAKNQG ISSVGRKLAE DRIVFSRHVL TAAAYVLGVH NRLHTGTYRL
 51 PSEVSAWDIL QKMRGGRPDS VTQIIEGSR FSHMRKVIDA TPDIGHDTKG
 101 WSNELKMAEV APDAFSGNPE GQFFPDSYEI DAGGSDLQIY QYAYKAMQRR
 151 LNEAWAGRD GLPYKNPYEM LIMASLIEKE TGHEADRDHV ASVFVNRLLFG
 201 GMRLQTDPSV IYMGAAAYKG KIRKADLRD TPYNTYTGGG LPFTRLALPG
 251 KAAMDAAAHF SGEKLYFVS KMDGTGLSQF SHDLTEHNAA VRKYILKK*

15 ORF7ng-1 and ORF7-1 show 98.0% identity in 298 aa overlap:

	10	20	30	40	50	60
orf7-1.pep	KLLKWSAVFLTVSAAVFAALLFVPKDNGRAYRKIAKNQGISSVGRKLAE DRIVFSRHVL					
orf7ng-1	YRIKIAKNQGISSVGRKLAE DRIVFSRHVL					
20	70	80	90	100	110	120
orf7-1.pep	TAAAYVLGVHNRHTGT YRLPSEVSAWDILQKMRGGRPDSVTQIIEGSRFSHMRKVIDA					
25	TAAAYVLGVHNRHTGT YRLPSEVSAWDILQKMRGGRPDSVTQIIEGSRFSHMRKVIDA					
30	130	140	150	160	170	180
orf7-1.pep	TPDIGHDTKGWSNEKLMAEVAPDAFSGNPEGQFFPDSYEIDAGGSDLQIYQYAYKAMQRR					
orf7ng-1	TPDIGHDTKGWSNEKLMAEVAPDAFSGNPEGQFFPDSYEIDAGGSDLQIYQYAYKAMQRR					
35	190	200	210	220	230	240
orf7-1.pep	LNEAWESRODGLPYKNPYEMLIMASLVEKETGHEADRDHVASFVNRLLKIGMRLQTDPSV					
orf7ng-1	LNEAWAGRDGLPYKNPYEMLIMASLIEKETGHEADRDHVASFVNRLLKIGMRLQTDPSV					
40	250	260	270	280	290	300
orf7-1.pep	IYMGAAAYKGKIRKADLRDTPYNTYTRGGLFPPTIALPGKAALDAAAHFSGEKLYFVS					
orf7ng-1	IYMGAAAYKGKIRKADLRDTPYNTYTGGLFPPTIALPGKAAMDAAAHFSGEKLYFVS					
45	310	320	330			
orf7-1.pep	KMDGTGLSQFSHDLTEHNAAVRKYILKKX					
orf7ng-1	KMDGTGLSQFSHDLTEHNAAVRKYILKKX					
50	280	290				

In addition, ORF7ng-1 shows significant homology with a hypothetical *E. coli* protein:

sp|P28306|YCEG_ECOLI HYPOTHETICAL 38.2 KD PROTEIN IN PABC-HOLB INTERGENIC REGION
 gi|1787339 (AE000210) o340; 100% identical to fragment YCEG_ECOLI SW: P28306 but
 has 97 additional C-terminal residues [Escherichia coli] Length = 340
 Score = 79 (36.2 bits), Expect = 5.0e-57, Sum P(2) = 5.0e-57
 Identities = 20/87 (22%), Positives = 40/87 (45%)

Query: 10 GISSVGRKLAE DRIVFSRHVL TAAAYVLGVHNRHTGT YRLPSEVSAWDILQKMRGGRPD 69
 G ++G +L D+I+ V + + + GTYR +++ ++L+ + G+
 Sbjct: 49 GRLLALGEQLYADKIIINRPRVFQWLLRIEPLDSHFKAGTYRFTPQMTVREMLKKLLSSEK 108

Query: 70 SVTVQIIEGSRFSHMRKVIDA TPDIGH 96

-90-

```

          ++++EG R S   K +   P I H
Sbjct:   109 QFFLRLEGMRLSDYLKQLREAFYIKH 135

    score = 438 (200.7 bits), Expect = 5.0e-57, Sum P(2) = 5.0e-57
    Identities = 84/155 (54%), Positives = 111/155 (71%)

5      Query:   120 EGGFFPDSYBIDAGGSDLQIYQTAYKAMQRRLINEAWAGRDGLPYKNPYEMLIMASLIEK 179
          EG F+PD++   A +D+ + + A+K M + ++ AW GR DGLPYK+ +++ MAS+IEK
Sbjct:   158 EGFWFPTDWMYTANTTDVALLKRAHKMKVKAVDSAWEGRADGLPYKDKNLVTMASIEK 217

10     Query:   180 ETGHEADRDHVASVFNRLKIGMRLQTDPSVIYGMGAAYGKIRKADLRDTPYNTYTG 239
          ET ++RD VASVF+NRL+IGMRLQTD+VIYGMG Y GK+ +ADL T YNTYT
Sbjct:   218 ETAVASERDKVASVFINRLRIGMRLQTDPTVIYGMGERYNGKLSRADLETPTAYNTYTI 277

15     Query:   240 GLPPTRIALPGKAAMDAAAHPSGEKYLYFVSKMDG 274
          GLPF IA PG ++ AAAHP+ YLYFV+ G
Sbjct:   278 GLPPGAATATPGADSLKAAAHFAKTPYLYFVADGKG 312

```

Based on this analysis, including the fact that the *H. influenzae* YCEG protein possesses a possible leader sequence, it is predicted that the proteins from *N. meningitidis* and *N. gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 6

The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 39>:

```

1      CGTTTCAAAA TGTTAACTGT GTTGACGGCA ACCTTGATGT CCGACACGGT
51     ATCTGCGGCC GGAAGCGCGTG CGGGGGATAT GAACACAGCGC AAGGAAGTCG
101    GAAAGGTTTT CAGAAAGCAG CAGCGTTACA CGCAGGAAGA AATCAAAAAC
151    GAACGCGCAC GGCTTGCGGC AGTGGCGCAG CCGGTTAATC AGATATTTAC
201    GTTGCTGGGA GGGGAAACCG CTTGCAAAA GGGGCAGCGC GGAACCGCTC
251    TGGCAACCTA TATGCTGATG TTGGAACGCA CAAATCCGCC CGAAGTCGCC
301    GAACGCGGCT TGAAGATGGC CGTGTCGCTG AACGCGTTTG AACAGCGGGA
351    AATGATTATC CAGAANTGGC GCGCATATGA GCTTATACCG GGTARAGGCG
401    AAAAACGGGC GGGGTGGCTG CGGAACGTGC TGAAGGACAG AGGAATCAG
451    CATCTGGACG GACGGGAAGA AGTGTGCTGC CAGCGGACG AAGGACAG

```

This corresponds to the amino acid sequence <SEQ ID 40; ORF9>:

```

1      ..RFKMLIVLTA TLIAGQVSAA GGGAGDMKQP KEVGKVFRRQ QRYSEEEKIN
51     ERARLAAVGE RVNQIFTLTG GETALQKGQA GTALATYMLM LERTKSPVEA
101    ERALEMAVSL NAFEQAEMIIY QKWRQIEPIP GKAKRAGWL RNVLRERNGQ
151    HLDGREEVLA QADEGQ

```

Further sequence analysis revealed the complete DNA sequence <SEQ ID 41>:

```

1      ATGTTACCTA ACOGTTTCAA AATGTTAACT GTGTGACGG CAACCTTGAT
51     TGCCGGACAG GATCTGCGCG CCGGAGGCGG TGCGGGGGAT ATGAAACAGC
101    CGAAGGAAGT CGGAAGGTTT TTCAAGAAAGC AGCAGCGGTA CAGCGAGGAA
151    GAATCACTAA ACGAACGCGC ACGGCTTGCG GCAAGTGGCG AGCGGGTTAA
201    TCAGATATTT ATGTTCTGCG GAGGGGAAC CCGCTTGGCA AAGGGGACAG
251    CCGGAAACGCG TCTGCAACCC TATATCCTCG TCTTGGCAAC CACAAATATC
301    CCGAAGTTCG CCGAAGCGCG CTTGGAATAT GCGGTGTGCG TGAACGCTT
351    TGAACAGGCG GAAATGATTT ATCAGAAATG GCGGCAGATT GAGCCTATAC
401    CCGGTAAAGC GCAAAAACGG CGGGGTGCGC TGCGGAACGT GCTGAGGGAA
451    AGAGGGAATC AGCATCTGGA CGGACTGGAA GAAGTGTCTG CTAGGCGGGA
501    CGAAGGACAG AACCGCAGGG TGTTTTTATT GTTGGCACAA GCOCGCGTGC
551    AACAGGACGG GTTGCGCGAA AAGCATCGGA AAGCGGTTGC CCGCGCGGCG
601    TTGAATATG AACTCTGCGC CGAAGCGGCG GTTGCAGATG TGGTGTTCAG
651    CGTACAGGGA CGCGAAAAGG AAAAGGCAAT CGGAGCTTGC CAGCGTTTGG
701    CGAAGCTCGA TACGGAATAA TCGGCCCCCA CTTTAATGAC GTTGCGTCTG
751    ACTGCACGCA AATATCCCGA AATACTCGAC GGCTTTTTCG AGCAGACAGA
801    CACCCAAAC CTTTGCAGCG TCTGCGAGGA ATGTGGAATT ATGAATCTGG
851    TTTCCCTGCA CAGGCTGGAT GATGCTATG CCGGTTTGA ACGTGTGTTG

```


101	AAGTCGGGAA	GTTTTCAGTA	AAGCAGCAGC	GTTCACAGCA	GGAGAAGATC
102	AAAAACGAC	CGCAGCGGCT	TGGCGCATGT	CGGACCGGG	TTAATCAGT
201	ATTACCTGT	CTGGGAGGAG	AAACCGGCTT	GCAAAAGGGG	CAGGCGGGAA
202	CGCTCTGGC	AACTATATG	CTGATGTGG	ACGACCAAA	ATCCCCGGA
301	CTGCGCGAC	CGCTCTGGA	AATGGCGCGT	TGCTGTAAGC	CGTTTGAACA
302	GGCGGAATG	ATTATACGA	AATGGCGCGA	GATGAGCGT	ATACCGGGTA
401	ACGACGAAAA	ACGGGCGGCT	TGGCTGCGTA	ACGTGCTGAG	GGAAAGAGG
402	AATCAGCAT	TAGACGCGG	GGAAGAAAGT	CGCTCTCAGG	CGCAGCAAGN
501	ACGAGACCGG	AGGGTGTGTT	TATTTGTGGC	ACAAGCGCGC	GTGCAACAGT
551	ACGGGTGGC	GCAAAAGAGT	TGCAAGCGCG	TTCGCGCGGG	GGCGTTGAGA
601	TATGAACAT	TGCCCGAAGC	GGCGTGTGCC	GATGTGGGTG	TCAGCGTACA
651	GGNACCGGAA	AGGAAAGAG	CAATCGGAGC	TTGTGCGGT	TCGCGAAGC
701	TCGATACGGA	AATATTGCC	CCCACTTAA	TGACGTTGCG	TCCTAGCTCA
751	CGCAAAATC	GCGAAATAT	CGACGGCTTT	TTTCAGACGA	CAGCACACCA
801	AARCTCTTG	CGCGTCTGCG	AGGAAATGGA	AATATGAT	CTGTTTCCC
851	TCGCAGGCT	GGATGATGCC	TATGCGGCT	TGAACCTGCT	GTGGAACGCG
901	AATCGAATG	CAGACCTGTA	TATCAGCGA	CGGATATTGG	CGGCAAAACG
951	AAAGAANGT	GCTTCGGTGA	TCAGCGCTGA	CGCGAAGAG	GCAATACGCA
1001	GGGGGACGCG	GGAACACGGG	GCGACGGCGG	CAATGACGCG	GCGCATAGAT
1002	TATCGCGACC	GAGGAGTATA	CACAAAGTC	AGCAGAGTGT	TGA AAAAGT
1101	CTCGCGCGCG	GATACCTGTT	TCGACAAGG	TGTCTGGGCG	GCTCGCGCGG
1151	CTGTGAGTT	GAATGCGGCG	AGGGCGGCT	TCGCGCAAGT	CGCAGCGSTG
1201	CGGAACCTC	CGGAACAGCA	GGGGCGGAT	TTTACGCGAG	ACAAATTGTC
1251	CAAAATCAG	ATTGTCCGCC	TGTCGATCT	CGCGGACAA	CGGGAGGCTT
1301	TGGCGGGGTT	GGACAGGATT	ATCGAAAAGC	CGCTCTCGGG	AGCATTAATCA
1351	GAGTTCAGC	GACGAGCAT	GGTACACGGG	TCAGTTTGTG	TCAGTGGGCT
1401	TGCGACATGA	AAAAAATAT	TTTTCAGATCT	TGAAGAGSGT	TTCCAGGCTT
1451	CCGCGATGA	CGCTCAGAT	ATGATATATC	TGGGCTACAG	CTCTGTTCC
1501	GATTCCGAT	TCGCTTACGA	ACGCTTCAGA	TCGCTCTCAG	CGCAGTACCA
1551	AATCAACCG	CGAGTACATG	TGCTCAGAG	CGGCGGTATC	CGGCGGTATC
1601	AACTGAAGG	CAGACGCGAA	ACGCGCTGCG	CGTATGTGCG	GATTCGCTT
1651	CAAAACACG	CGAGCGCCGA	AGTTGCGCGT	CGTTTGGGCG	AACTGCTT
1701	GGCAATTGGG	CAGACGCGAT	AGGGGSGTGA	CGTATGGAGC	CAGCGCGCAC
1751	ACCTTACGCG	AGACAGAA	ATATGGCGGAG	AAAGCTCTAA	ACGTACGCGC
1801	ATCGCATTCG	CCCAACTCT	CCGAAAGACT	CGGAATCBA	

This encodes a protein having amino acid sequence <SEO ID 44>.

1	MLEPAPFTILS	VLAALAAFLAQ	AYGAAGAADAK	PPKEVGKATVF	KQQRYSSEET
5	KNERARLAPAV	GERVQVIFQIT	LYXETALAGATV	LMERTKSPER	
101	VNARALEMAV	SNIAFEQAMR	YQVWKQRIEF	IPGKQAKRAG	WLRNVLRERG
151	QOHLDGLEEX	LQADEQOQNR	RVFLLLAQAA	VQODGLQJKA	SKAVRRALRL
201	YEHLEPAAVA	DVFSVSGORR	KEKATGALGR	LAKLTDLTP	PTMTLRLTA
251	RKYPEILDFG	FEGTQDQNL	AWQBMESIM	LVSJLRHDL	YARLNVLLER
301	NPMDALQQA	AILAANRKEK	ASYVDGYAEK	AYGRGTGEGR	GRAMTAMAI
351	YADRDRYTVK	RQWLKVSQK	EYFDPKGLVA	ANAVALDXGR	RNALRQIGRV
401	RKLPEQOGVR	FADNLNLSK	MFALSKIPOL	REALRQDQKI	ITEPKGASNP
451	ELQASIVRGR	SVYDNLGRK	KMI3LDERA	FLRAPLNDQI	MNNGLQ5LSL
501	DSKRIKQVLA	LICRTAYQIN	EDQAVNSDST	WAYYLKQDTE	SALPYLRYSF
551	ENDEPPEVAJ	LELVGEVWALP	DRDQAVDVMT	WAQXLHDKKK	IWRETLKRGH
601	IAPLOPSRKP	RK*			

ORF9a and ORF9-1 show 95.3% identity in 614 aa overlap:

[illegible]

	orf9-1	EMIQKWRQIEPIGKAQKRAGLRLNVLRRGNQHLDGLLEVLQAQDEGQNRVRVLLLAQ	130	140	150	160	170	180
5	orf9a.pep	AAVQQDGLAQKASKAVRRAALRYEHLPEAAVADVVSQVQREKEKAIQALQRLAKLDTEI	180	190	200	210	220	230
	orf9-1	AAVQQDGLAQKASKAVRRAALKYEHLEPEAAVADVVSQVQREKEKAIQALQRLAKLDTEI	190	200	210	220	230	240
10	orf9a.pep	LPPTLMTLRITARKYPEILDGFFEQTDTQNLISAVWQEMEIMNLVSLHRLDDAYARLNVLL	240	250	260	270	280	290
	orf9-1	LPPTLMTLRITARKYPEILDGFFEQTDTQNLISAVWQEMEIMNLVSLHRLDDAYARLNVLL	250	260	270	280	290	300
15	orf9a.pep	ERNPNADLYIQAAILAANKKEXASVIDGYAEKAYGRGTGEQGRGAAMTAAMIYADRRDYT	300	310	320	330	340	350
20	orf9-1	ERNPNADLYIQAAILAANKKEGASVIDGYAEKAYGRGTGEQGRRAALTAMMIYADRRDTA	310	320	330	340	350	360
	orf9a.pep	KVRQWLKKVSAPFYLFDKGVLAATAAAVELDXGRAALRQIGVRVKLPEQQGRYFTADNLSK	360	370	380	390	400	410
25	orf9-1	KVRQWLKKVSAPFYLFDKGVLAATAAAVELDGRAALRQIGVRVKLPEQQGRYFTADNLSK	370	380	390	400	410	420
30	orf9a.pep	IQMFALSKLPDKREALRGLDKIIEKPPAGSNTELQAEALVQRSVVYDRLGKRKKMISDLE	420	430	440	450	460	470
	orf9-1	IQMLALSKLPDKREALRGLDKIIEKPPAGSNTELQAEALVQRSVVYDRLGKRKKMISDLE	430	440	450	460	470	480
35	orf9a.pep	RAFRLAPDNAQIMNNLGYSLTDSKRLDEGFALLQTAYQINPDDTAVNDISIGWAYYLKXD	480	490	500	510	520	530
	orf9-1	RAFRLAPDNAQIMNNLGYSLTDSKRLDEGFALLQTAYQINPDDTAVNDISIGWAYYLKGD	490	500	510	520	530	540
40	orf9a.pep	AESALPYLRYSFENDPEPEVAHLEGEVLWALGERDQADVVTQAAHLTGDDKKIWRETLKR	540	550	560	570	580	590
	orf9-1	AESALPYLRYSFENDPEPEVAHLEGEVLWALGERDQADVVTQAAHLTGDDKKIWRETLKR	550	560	570	580	590	600
45	orf9a.pep	HGIALPQPSRKPRKX	600	610				
50	orf9-1	HGIALPQPSRKPRKX	610					

Homology with a predicted ORF from *N.gonorrhoeae*

ORF9 shows 82.8% identity over a 163aa overlap with a predicted ORF (ORF9.ng) from *N.*

gonorrhoeae:

55	orf9	RFKMLT ¹ VL ² AT ³ LIAG ⁴ QV ⁵ SAAGGAGD ⁶ MKQ ⁷ PEV ⁸ GV ⁹ K ¹⁰ FK ¹¹ QQ ¹² Q ¹³ Y ¹⁴ SE ¹⁵ E ¹⁶ E ¹⁷ K ¹⁸ NER ¹⁹ AR ²⁰	54
	orf9ng	MIML ¹ PAR ² T ³ FL ⁴ SV ⁵ LAAL ⁶ LAG ⁷ Q ⁸ Y ⁹ AA ¹⁰ --GA ¹¹ VD ¹² EL ¹³ P ¹⁴ KEV ¹⁵ GV ¹⁶ K ¹⁷ LKH ¹⁸ RY ¹⁹ SE ²⁰ E ²¹ E ²² K ²³ NER ²⁴ AR ²⁵	58
60	orf9	LA ¹ AV ² GER ³ VN ⁴ Q ⁵ I ⁶ FT ⁷ LL ⁸ GG ⁹ ET ¹⁰ AL ¹¹ Q ¹² K ¹³ Q ¹⁴ AG ¹⁵ T ¹⁶ AL ¹⁷ AT ¹⁸ Y ¹⁹ ML ²⁰ LM ²¹ ERT ²² K ²³ SP ²⁴ EA ²⁵ RA ²⁶ LEM ²⁷ AV ²⁸ SL ²⁹ NA ³⁰ FE ³¹	114
	orf9ng	LA ¹ AV ² GER ³ VN ⁴ V ⁵ FT ⁶ LL ⁷ GG ⁸ ET ⁹ AL ¹⁰ Q ¹¹ K ¹² Q ¹³ AG ¹⁴ T ¹⁵ AL ¹⁶ AT ¹⁷ Y ¹⁸ ML ¹⁹ LM ²⁰ ERT ²¹ K ²² SP ²³ EA ²⁴ RA ²⁵ LEM ²⁶ AV ²⁷ SL ²⁸ NA ²⁹ FE ³⁰	116
	orf9	Q ¹ REMI ² Y ³ Q ⁴ K ⁵ W ⁶ Q ⁷ I ⁸ E ⁹ PT ¹⁰ PG ¹¹ AQ ¹² K ¹³ AG ¹⁴ LV ¹⁵ LN ¹⁶ VR ¹⁷ ER ¹⁸ GN ¹⁹ Q ²⁰ HL ²¹ D ²² GE ²³ VE ²⁴ VL ²⁵ QA ²⁶ DE ²⁷ Q ²⁸	168
65	orf9ng	Q ¹ REMI ² Y ³ Q ⁴ K ⁵ W ⁶ Q ⁷ I ⁸ E ⁹ PT ¹⁰ PG ¹¹ EA ¹² Q ¹³ K ¹⁴ PA ¹⁵ GV ¹⁶ LV ¹⁷ LN ¹⁸ VR ¹⁹ KE ²⁰ GG ²¹ N ²² PH ²³ LD ²⁴ RE ²⁵ EV ²⁶ PA ²⁷ QS ²⁸ DY ²⁹ VH ³⁰ Q ³¹ PM ³² I ³³ FL ³⁴ LL ³⁵	178

The ORF9ng nucleotide sequence <SEQ ID 45> was predicted to encode a protein having including acid sequence <SEQ ID 46>:

```

1  MIMLPAREFTT LSVLAAALLA GQAYAGAAD VELPKVEVGKV LRKHRRYSEE
51 EIKNERARLA AVGERVNRVF TLLGGETALQ KGQAGTALAT YMLLERTKTS
101 PEVAERALEM AVSLNAFEQA EMIYQKWRQI EPIPGEAQKP AGWLNRNLKE
151 GGNPHLDRL E VPAQSDYVH QPMIFLLVQ AAQVHGGSVAQ KPSKAVRPA
201 YNVEVLPETA GADAVFCVQG PQYEKAIQSF PFCGRNPQTE NIAPFFNELF
251 RPTARFISPK LLQRFRTETP NLAKFRFPFG PEMETYQTGE ERLFLTRNPT

```

Amino acids 1-28 are a putative leader sequence, and 173-189 are predicted to be a transmembrane domain.

Further sequence analysis revealed the complete length ORF9ng DNA sequence <SEQ ID 47>:

```

1  ATGTTACCCG CCGCTTTCAC TATTTTATCT GTCCCTGCGAG CAGCCCTGCT
51 TGGCCGGACAG GCGTATGCTG CCGGCGCGGC GSGATGTGAGG CTGCCGAGAGG
101 AAGTCGGAA A GGTTTTAAAG AAACATCGCG GTTACAGCGA GSRAGAATATC
151 AAAAAACGAAC GCGCACGGCT TCGCGCAGTG GCGCAACGGT TCACACAGGT
201 GTTTACGGCTG TCGCGCGGTG AAACGCGCTT GCAGCAAGGG CAGCGGGGAA
251 CGCCTCTGGC AACCTATATG CTGATGCTGG AACGCACAAA ATCCCCCGAA
301 CTGCGCGAAGC GCGCCTTGGA AATGCGCTGT TCCCTGAACG GCTTGAACA
351 GCGCGAAATG ATTTATCAGA AATGcgcgca gatcgagcct ataCcggggt
401 agggcgcaaaa accgGggggG tggctgcgga acgtattgaa ggaagggGga
451 aaTCAGCATC TGGAcggggt gaaagaggTG CtggcgcaAT cgagcgatGT
501 GCAAAAACgc aggaTATTTT TGCTGCTGGT CAAAGCCGCC GTGcagcagg
551 gTGGGGTGGC TCAAAAAGCA TCGAAAGCGG TTCGCGctgG GgctgtgaAG
601 TATGAACATC TGCCcgaagc ggcggTTGCC GATGcggTGT TCGCGGTACA
651 GGGACGCGAA AAGGAAAagg caaTCGAAGC TTTGCAAGCT TTGCGGAAGC
701 TCGGATACGGA AATATTGCC CCACTTTAA TGACGTTGCG TCTGACTGCA
751 CGCAAAATATC CCGAAATACT CGACGCGCTT TTGAGCAGA CAGACACCCA
801 ABAACCTTTCG GCGCTCTGCC AGGAAATGGA AATTATGAAT CTGTTTTCCT
851 TGGTAAGGCC GGATGATGCC TATGCGCGTT TGAACGTGCT GTTGAACAC
901 AACCCGAATG CAACACCTGTA TATTCAAGGG CCGCATATTTG CCGCAACCG
951 AAAAGAAGCT CGTCTCGTTA TCGACGCGTA GCGCGAAAG GGTACGGCA
1001 GGGGACCGG GGAACAGCGG GGCAGggcgG AATgacggc GCCGATGATA
1051 TATGCCGACC CAGGGGATTA CGCCAAATCT AGGCAGTGGT TGA AAAAGT
1101 GTCCGCGCGG GAATACCTGT TCGACAAGAG CGTGCTGGCG GCTGCGCGGG
1151 CTGCGCAATT GGACGGGAGC CGGCGGCTT TCGCGCAGAT CGGCAGGGTG
1201 CGGAATCTTC CCGAACAGCA GGGCGGATAT TTTACGGCAG ACAATTGTTC
1251 CAAAATACAG ATGCTCGCCC TGTCGAAGCT GCCCGACAAA CGGAAGGCC
1301 TGATCGGGCT GAACAACATC ATCGCCAAAC TTTGCGCGGC GGAAGCAGC
1351 GAACCTTTTG CGGAAGCATT GCACACAGCT TCCATTATTT ACgaaCGATT
1401 cggCAACCGG GGA AAAATGA TTGCGGACCT tgaAACcgcg CTCAACTTA
1451 CGCCCGGATA TGCACAAATT ATGAATAATC TGGGCTACAG CCTGCTTTCC
1501 GATTCCAAAC GTTTGGACGA GGGTTTCGCC CTGCTTCAGA CGGCATACCA
1551 AATCAACCCG GACGATACCG CGSTTAAOCA CAGCATAGGC TGGCGGTATT
1601 ACCTGAAAGG CGAGcgggaa AGCGCGCTGC CGTATTGCT gcatctcgtt
1651 gAAACGAGC CGAGACCGGA AGTTCGCGC CATTTGGCG AATGTTTGG
1701 GGCATGGGC GRACCGATC AGGCGGTTGA CGTATGGAG CAGCGGCAC
1751 ACCTTAGGGG AGACAGAGAA ATATGGCGGG AGACGCTCAA ACCTACGGGA
1801 ATCGCCTTGC CGAGCCTTC CGAAAACCC CGGAATAA

```

This encodes a protein having amino acid sequence <SEQ ID 48>:

```

1  MLPARTTILS VLAALLAGO AYAAGAADV LPEKVGKVLV KHRRYSEEEI
51 KNERARLA AVGERVNRVFTL LGGETALQKG QAGTALATYM LMLERTKSEF
101 VAERALEMAV SLNAFEQAE EMIYQKWRQIE IPGEAQKPAG WLRNLKEGG
151 NQHLDLGKEV LAQSDVQKR RIFLLVQAA VQGGVQAQKA SKAVRRALK
201 YELHPEAAVA DAVFGVQGRE KEKAIEALQR LAALDELTLR PTMLTRLRLA
251 RKYFEILDGF FEQTDQNL S AVWQEMEIMN LVSLRPDDPA YARLNVLLHE
301 NPNANLYQIA ALLAA NRKEG ASVIDGYAEK AYRGRTGEQR GRAAMTAAMI
351 YADRDRYAKV RQWLKRV SAP EYLFDRGVIA AAAAELDGG RAALRQIGRV
401 RKLFEQGGRY FTADNL SKIQ MLALS KLPDK REALIGLNNI IAKLSAAGST
451 EPLAEALQR SIIEQFGKR GMLADLETA LKLTFPDQAI MNNGYSLLS

```


501 DSKRLDEGFA LLQTAYQINF DDTAVNSIG WAYYLLKGD AE SALPYLRYSF
 551 ENDEPEVA AHLEVLNAG ERDQAVDVWT QAAHLRGDKK IWRETLKRYG
 601 IALPEPSRKF RK*

ORF9ng and ORF9-1 show 88.1% identity in 614 aa overlap:

		10	20	30	40	50	60
5	orf9-1.pep	MLPNRFKMLT	VLTTATL	LAGVSAAGG	GAGDMKQ	PKVEGKVF	RKQORYS
	orf9ng-1	MLPARFT	ILSVLAALL	AGQAYAA	GAAG--	AADVLPKE	VGKVLKRRYS
		10	20	30	40	50	
10	orf9-1.pep	AVGERV	NQIF	FTLLG	GETALQ	KQAGTAL	ATYMLML
	orf9ng-1	AVGERV	NRVFT	LLGGET	ALQKQAG	TALATY	MLMLERT
		60	70	80	90	100	110
15	orf9-1.pep	EMIIQK	WRQIE	PIPGK	AQRAG	WLRN	VLRE
	orf9ng-1	EMIIQK	WRQIE	PIPGK	AQRAG	WLRN	VLRE
		120	130	140	150	160	170
20	orf9-1.pep	AAVQ	QDGLA	QKASKA	VRRAAL	KYEH	LP
	orf9ng-1	AAVQ	QDGLA	QKASKA	VRRAAL	KYEH	LP
		180	190	200	210	220	230
25	orf9-1.pep	LPPT	MTLRL	TARKY	PEIL	DGFF	EQ
	orf9ng-1	LPPT	MTLRL	TARKY	PEIL	DGFF	EQ
		240	250	260	270	280	290
30	orf9-1.pep	ERNP	NADLY	IQAA	ILAA	NRKE	GA
	orf9ng-1	ERNP	NANLY	IQAA	ILAA	NRKE	GA
		300	310	320	330	340	350
35	orf9-1.pep	KVRQ	WLKKV	SAPE	YLF	FDK	GV
	orf9ng-1	KVRQ	WLKKV	SAPE	YLF	FDK	GV
		360	370	380	390	400	410
40	orf9-1.pep	IQML	LSKLP	DKRE	AL	RLG	L
	orf9ng-1	IQML	LSKLP	DKRE	AL	RLG	L
		420	430	440	450	460	470
45	orf9-1.pep	RAFR	LAPD	NAQIM	NNLG	YSLL	TS
	orf9ng-1	RAFR	LAPD	NAQIM	NNLG	YSLL	TS
		480	490	500	510	520	530
50	orf9-1.pep	AESAL	PYLRY	SFEND	PEVA	AHL	GE
	orf9ng-1	AESAL	PYLRY	SFEND	PEVA	AHL	GE
		540	550	560	570	580	590
55	orf9-1.pep	HGIAL	PQPS	RKPR	KX		
	orf9ng-1	HGIAL	PEPS	RKPR	KX		
		600	610				
60	orf9-1.pep						
	orf9ng-1						
65	orf9-1.pep						
	orf9ng-1						

600 610

In addition, ORF9ng shows significant homology with a hypothetical protein from *P.aeruginosa*:

```

sp|P42810|YHE3_PSEAE HYPOTHETICAL 64.8 KD PROTEIN IN HEMM-HEMA INTERGENIC REGION
(ORF3)
>gi|1072999|pir||s49376 hypothetical protein 3 - Pseudomonas aeruginosa >gi|557259
(X82071) orf3 [Pseudomonas aeruginosa] Length = 576
Score = 128 bits (318), Expect = 1e-28
Identities = 138/587 (23%), Positives = 228/587 (38%), Gaps = 125/587 (21%)

10 Query: 67 VFTLLGGETALQKQAGTALATYMLMLERTKSPVEAERALEMAVSLNAFEQAEMIYQKWR 126
+++LL E A Q+ + AL+ Y++ ++T+ P V+ERA +A L A ++A W
Sbjct: 53 LYSLLVAELAGQRNRFDIASNYVVOAQKTRDPGVSEAFRTAEYLGDAQEALDTSLLWA 112

15 Query: 127 QIEPIPGAQKFPAG-----WLRNVLKEGGNQHL DGLKEVLAQSDDVQKRR 172
+ P +AQ+ A ++ VL G+ H D L A++D + +
Sbjct: 113 RSAPDNLDQRAAAIQLARAGRYEESMVYMEKVLNGQGDTHFDLALSAAETDPDTRAGL 172

20 Query: 173 FXXXXXXXXXXXXXKASKAVERRAALKYEHLPAAVADAVFGVQGREKEKAIEALQRLA 232
+++++ KY + + A+ Q ++A+ L+ +
Sbjct: 173 L-----QSFHLLKKYPNNQGLLFGKALLQQDGRPFDEALTLELNS 214

25 Query: 233 KLDTEILPPTIMTLRLTARK----YPEILDGFEQTDTQNL SAVWGEIMNLVSLRKP 287
E+ P L+ L+ K P+ G E D+ + + + LV+
Sbjct: 215 ASRHEVAPLLRLSRLLSQMKRSDEALPLLKAGKEHPDDKRVRLAYARI----LVEQNR 270

30 Query: 288 DDAYARLNVILLEHNP-----ANLYTQAAT----- 312
DDA A L++ P+ A+Y++ +
Sbjct: 271 DDAKAEFAGLVQFPDDDDDLRFSLALVCLAEQAWDEARIYLEELVERDSHVDAAHFNLG 330

35 Query: 313 -LAANKEGASVIDGYAEKAYGRGTGEQGRAAMTAAMIYADRRDYAKRVQWLKQVSAPE 371
LA +K+ A +D YA+ G G + T ++ A R D A + + P+
Sbjct: 331 RLAEQKDTARALDEYAQ--VGFNDLFLAQLRQTDVILKAGRVDEAAQLRDKARSEQ 388

40 Query: 372 YLFDKXXXXXXXXXXXXXXXXXRXRQIGRVKLPQGGRYFTADNLSKIQMLASKLDPKR 431
Y A L T+ ALS +
Sbjct: 389 Y-----AIQLYLI EAELSNNDDQE 408

45 Query: 432 EALIGLNNIIAKLSAAGSTEPLAEALAQRSIIYEQFGKRGKMIADLETALKLTPDNAQIM 491
+A + + + E L L RS+ E+ +M DL + PDNA +
Sbjct: 409 KAWQAIQEGLKQYP----EDL-NLTYTRSLAEKRNLDLQMEKDLRFVIAEPDNAMAL 462

50 Query: 492 NNLGYSLLSDSKRLDEGFALLQTAYQINPDDTAVNDSIGWAYYLKGDAESALPYRYSFE 551
N LG+L + R E L+ A++NPD A+ DS+GW Y+ G A YLR + +
Sbjct: 463 NALGTYLADRTTRYGEARELILKAHLKNPDDPAIDLSMGWYINRYQSKGLADAEYRLQALQ 522

55 Query: 552 NDPEPEVA AHLGEVLWALGERDQAVDVTQAHLRGDKKIWRETLKR 598
P+ EVA AHLGEVLWA G + A +W + + + D + R T+KR
Sbjct: 523 RYPDHEVA AHLGEVLWAGRQGDARAIWREYLDKQPDSDVLARTIKR 569

60 Query: 553 gi|2983399 (AE000710) hypothetical protein [Aquifex aeolicus] Length = 545
Score = 81.5 bits (198), Expect = 1e-14
Identities = 61/198 (30%), Positives = 98/198 (48%), Gaps = 19/198 (9%)

65 Query: 408 GRYFTADNL-SKIQMLASKLDPKREALIGLNNIIAKLSAAGSTEPLAEALQ----- 459
G Y A L L K ++LA PDK+E L + +K + + + L+
Sbjct: 335 GNYEDAKRLIEKAVLA----PDKKEILFEADYYSKTKQYDKALEILKLEKDYDPNDR 390

Query: 460 ----RSIYEQFGKRGKMIADLETALKLTPDNAQIMNNLCYSLLS--DSKRLDEGFALLQ 513
++Y+ G G L A++L P+N N LCVSL +R++E L++
Sbjct: 391 VYFMEAIVDNLGDIKNAEKALRKAEIDPENPDYNNLYGYSLLWYKGERVEAEELIK 450

Query: 514 TAYQINPDDTAVNDSIGWAYYLKGDAESALPYLRYSF-ENDPEPEVA AHLGEVLWALGER 572
A + +P++ A DS+GW YYLKG D A+ YL + E +P V H+G+VL +G+
Sbjct: 451 KALEKDPENPAYIDSMGWVYYLKG DYERAMQYLLKALREAYDDPVEVNEHVGVDLLKMGYK 510

Query: 573 DQAVDVWQAHLRGDKK 590
++A + + +A L + K
Sbjct: 511 EEARNYERALKLLEEGK 528

```

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 7

- 5 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 49>:

```

1 AACCTCTACG CCGGCCCGCA GACCACATCC GTCATCGCAA ACATCGCCGA
51 CAACCTGCAA CTGGCCAAAG ACTACGGCAA AGTACATGG TTCGCTCGCC
101 CGCTCTTCTG GCTCCTGAAC CAACTGCACA ACATCATCGG CAACTGGGGC
151 TGGGGGATTA TCGTTTAAAC CATCATGCTC AAAGCCGTAC TGTATCCATT
201 GACCAACGCC TCTTACCGCT CTATGGCGAA AATGCGTGCC GCCGACCCA
251 AACTGCAAGC CATCAAGAGG AATACGGCGG ACGACCGTAT GCGCAACCAA
301 CAGGCGATGA TGCAGCTTTA CACAGACGAG AAAATCAACC CGACTGGGCG
351 GGTGCTGCGC TATCGTGTGG CAAATCCCGG TCTTCATCGG ATTGTATTGG
401 GCATGTTTGG CTTCCGTAGA ATTGGCGCAG GCACCTTGGC TGGGTGTGAT
451 TACCGACCTC AGCGCGCGCG ACCCTACTTA CATCTGCGCC ATCATATTGG
501 CGCGAAGCCT CCGGACCGCA TTTATCTCGA ACCCGCGCGC GACGACCGCG
551 ATGCAAGCGA AATGATGAA AATCATGCGG TTGGTTTCTT CAGWCTTGTT
601 CTTCCTTCTC CTTGCGCGGk TGGTATTGTA CTGGTGTAGT AACACCTTCC
651 TGCACATCGC CCAGCAATGG CACATCAACC GCAGCATCGA AAAACACGCG
701 GCCCAAGGCG AAGTCGTTTC CTA

```

This corresponds to the amino acid sequence <SEQ ID 50; ORF11>:

```

1 ..NLYAGPQTTS VIANIADNLQ LAKDYGVVH FASPLFWLLN QLNHIINWGW
51 WAIIVLTIIV KAVLYPLINA SYRSMAMRA AAPKLOAIKE KYGDDRMARQQ
101 QAMMLYLTDE KINFLGGCLP MLLQIPVFIG LYWALFASVE LRQAPLWGI
151 TDLSRADPYF ILPIIMAART MAQTYLNPPP TDEMQAMMK IMPLVLSXXE
201 FFFPAGXVLY WVVNNLLTIA QQWHINRSIE KQRAQGEVVS *

```

Further sequence analysis revealed the complete DNA sequence <SEQ ID 51>:

```

1 ATGGATTTTA AAAGACTCAC GCGTTTTTTC GCCATCGCGC TGGTGATTAT
51 GATCGGCTGG GAAAAGATGT TCCCACTCC GAGGCAGTGC CCGCGCGCCC
101 AACAGGCGAGC ACAACACAGC GCGGTAAACCG CTTCGCGCGA AGCGCGCTCT
151 CGCGCGCGCAA CGCCGATTAC GGTARCGACC GACRACGTTC AGCCGTCTAT
201 TGATGAAARA AGCGCGGACG TGCGCGCGCT GACCGTGTCT AAATCAARA
251 CAACGCGCGA CGAAATARA CCGTCTATCC TGGTTGGCGA CGGCAAGAA
301 TACACCTACG TCGCCCAATC GCAACTTTTG GACGCGCAGG GCAACAGAT
351 TCTAAAAGGC ATCGGCTTTA GCGCACCGAA AAAACAGTAC AGCTTGGARG
401 GCGACAAAGT TGAAGTCGCG CTGAGCGCGC CTGAACACG CGGTCTGAAA
451 ATCGACAAGG TTTATACTTT CACCAAGGCG AGCTATCTGG TCAACGTCGG
501 CTTTCGACATC GCCAACGGCA CGGCTCAAC GCGCAACCTG AGCGCGGACT
551 ACCGCGATCGT CCGCGACCA CAGGGAACCG AGGGTCAAGG TTACTTTACC
601 CACTCTTACG TCGGCCCTGT TGTTTATACC CTTGAAGGCA ACTTCCAAAA
651 AGTCAGCTTT TCCGACTTGG ACGACGATGC CAAATCCGGC AAATCCGAGG
701 CCGAATACAT CCGCAAAACC CGACCGGCT GCGTCGGCAT GATTGAACAC
751 CACTTCATGT CCACTTGGAT TCTCAACCTT AAGGCGCAGC AAGCGGTTTG
801 CGCGCGCAGG GAGTGCAACA GCGACATCAA ACGCGCAAC GACAGGCTGT
851 ACAGCACGAG CGTCAGCGTG CTTTATACCG CCATCGAAAA CGCGCGAGTA
901 CGCGAGCCTT CATTACAAC CTACCGCGCG CCGCAGACA CATCCGTAT
951 CGCAACATCC GCGCAACACC TGCAACTGGC CAAAGACTAC GCGAAGTAC
1001 ACTGGTTCGC CTCGCCGCTC TTCTGGCTCC TGAACCACT GCACAACATC
1051 ATCGGCAACT GGGGCTGGGC GATTATCGTT TTAACCATCA TCGTCAAAGC
1101 CGTACTGTAT CCAATTGACA ACGGCTTTA CCGCTCATG GCGAAATGC
1151 GTGCGCGCGC ACCCAAACTG CAAGCATCA AAGAGAAATA CGGCGACGAC
1201 CGTATGGCGC AACACAGGCG GATGATGCG CTTTACACAG ACGAGAAAT
1251 CAACCCGCTG GCGGCTGCG TGCCTATGCT GTTGCAATC CCGCTCTCA
1301 TCGGATTGTA TTGGGCAATT TCGGCTCCG TAGAATTGCG CCAGGCACT
1351 TGGCTGGGTT GGATTACCGA CTTACGCGCG CCGGACCCCT ACTCATCTCT
1401 GCCCATCATT ATGCGGCGAA CGATGTTGCG CCAACTTAT CTGAACCCGCT
1451 CGCGACCGA CCGAATGCG GCGAAATCAT TGAAATCAT GCGGTTGGTT
1501 TTCTCGGCTA TGTCTTCTT CTCCCTGCC GGTCTGGTAT TGTACTGGTT

```

1551 AGTCAACAAC CTCCTGACCA TCGCCACGCA ATGGCACATC AACCGCAGCA
1601 TCGAAAACCA ACGCGCCCAA GCGAAGTCG TTCCTAA

This corresponds to the amino acid sequence <SEQ ID 52; ORF11-1>:

5 1 MDFKRLTAFF AIALVIMIGW EKMFFTPKPV PAPQQAQQQ AVTASAEAAAL
51 APATPITVTT DTVQAVIDEK SGDRLRLTIL KYKATGDENK PFILFGDGKE
101 YTYVAQSELL DAQNNILKG IGFSAPKKQY SLEGDKVEVR LSAPETRGLK
151 IDKYVTFKKG SYLVNVRFDI ANGSGQTANL SADYRIVRDH SEPEGQGYFT
201 HSYVGPVYVT PEGNFQKVSF SDLDDDAKSG KSEAEYIRKT PTGWLGMIEH
251 HFMSTWILQP KGRQSVCAAG ECNIDIKRRN DKLYSTVSUV PLAAIQNGAK
10 301 AEASINLYAG PQTTSVIANI ADNLQAKDY GKVHVFASPL FWLLNQLHNI
351 IGNWGWAIIV LTIIVKAVLY PLTNASYSRM AKMRAAAPKL QAIKEYGDD
401 RMAQQQAMMQ LYTDEKINPL GGCLPMLLIQ PVFIGLYWAL FASVELRQAP
451 WLGWITDLSR ADPYIILPII MAATMFAQTY LNPPPTDPMQ AKMMKIMPLV
501 FSVMFFFFPA GLVLYWVNN LTIAQQWHI NRSIEKQRAQ GEVVS*

15 Computer analysis of this amino acid sequence gave the following results:

Homology with a 60kDa inner-membrane protein (accession P25754) of *Pseudomonas putida*

ORF11 and the 60kDa protein show 58% aa identity in 229 aa overlap (BLASTp).

20 ORF11 2 LYAGPQTTSVIANIADNLQAKDYGKVHVFASPLFWLLNQLHNIIGNWGWAIIVLTIIVK 61
LYAG+ S + ++ L+L DYG + + A P+FWLL +H+++GNWGW+IIVLTI++K
60K 324 LYAGPKQSKLKEPLSGLELTVDYGFLWFIQAQPIFWLLQHIHSLNGNWCIIIVLTMLIK 383
ORF11 62 AVLYPLTNASYSRMAKMAAAAPKLQAIKEYGDDRRXXXXXXXXLYTDEKINPLGGCLPM 121
+ +PL+ ASYRSMA+MRA APKL A+KE++GDDR LY EKNINPLGGCLP+
25 60K 384 GLFFFLSAASYSRMARMRAVAPKLAALKERFGDDRQKMSQAMMELYKKEKINPLGGCLPI 443
ORF11 122 LQIPVPFIGLYWALFASVELRQAPWLGWITDLSRADPYIILPIIIMAAATMFAQTYLNPPPT 181
L+Q+PVE+ LYW L SVE+RQAPW+ WITDLS DP++LLEIM ATMF Q LNP P
30 60K 444 LVQMPVFLALYVWVLESVMRQAPWILWITDLSIKDPFFILPIIMGATNFIQQRNPPTFP 503
ORF11 182 DPMQAKMKIMPLVXXXXXXXXXPAQKVLVWVNNLTIAQQWHINRSIE 230
DPMQAK+MK+MP++ PAG VLYWVNN L+I+QW+I R IE
60K 504 DPMQAKVMKMPPIIFTFFFLWFPAGLVLYWVNNCLISQWYITRRIE 552

Homology with a predicted ORF from *N.meningitidis* (strain A)

35 ORF11 shows 97.9% identity over a 240aa overlap with an ORF (ORF11a) from strain A of *N. meningitidis*:

orf11.pep NLYAGPQTTSVIANIADNLQAKDYGKVHW
40 orf11a IKRRNDKLYSTSVSVPLAAIQNGAKSXASINLYAGPQTTSVIANIADNLQXKDYGKVHW
280 290 300 310 320 330
orf11.pep 40 50 60 70 80 90
FASPLFWLLNQLHNIIGNWGWAIIVLTIIVKAVLYPLTNASYSRMAKMAAAAPKLQAIKE
45 orf11a FASPLFWLLNQLHNIIGNWGWAIIVLTIIVKAVLYPLTNASYSRMAKMAAAAPKLQAIKE
340 350 360 370 380 390
orf11.pep 100 110 120 130 140 150
KYGDDRMQAQQAMMQLYTDEKINPLGGCLPMLLIQIPVFIGLYWALFASVELRQAPWLGI
50 orf11a KYGDDRMQAQQAMMQLYTDEKINPLGGCLPMLLIQIPVFIGLYWALFASVELRQAPWLGI
400 410 420 430 440 450
orf11.pep 160 170 180 190 200 210
TDLRADPYIILPIIIMAAATMFAQTYLNPPPTDPMQAKMKIMPLVFSXKFFFPAGKVLVY
55 orf11a TDLRADPYIILPIIIMAAATMFAQTYLNPPPTDPMQAKMKIMPLVFSXKFFFPAGKVLVY
460 470 480 490 500 510

orf11.pep 220 230 240
 WVNNLLTIAQWHINRSIEKQRAQGEVVSX
 ||:|||||
 5 orf11a WVNNLLTIAQWHINRSIEKQRAQGEVVSX
 520 530 540

The complete length ORF11a nucleotide sequence <SEQ ID 53> is:

1 ANGGATTTTA AAGACTCAC NNGTTTTTTC GCCATCGCAC TGGTGATTAT
 51 GATCGGATNG NAAANGATGT TCCCCACTCC GAAGCCCGTC CCGCGCGCCC
 101 AACAGACGGC ACAACAACAG GCCGTAAACG CTTCCGCGGA AGCCGCGCTC
 151 GCGCCCGNAN CGCCGATTAC CGTAACGACC GACACGGTTC AAGCCGCGAT
 201 TGATGAAAAA AGCGCGGACC TCGCGCGGCT GACCCTGCTC AAATACAAAG
 251 CAACCGGCGA CNAAAATAAA CCGTTCATCC TGTTTGGCGA CGGCAAAANA
 301 TACACCTACN TCGCCCAANTC CGAACTTTTG GAGCGCGAGG GCAACAACAT
 351 TCTAAAAGGC ATCGGCTTTA GCGACCGGAA AAACAGTACG AGCTTGGRAG
 401 GCGACARAAG TGAAGTCCGC CTGAGCGCAC CTGAACAACG CGGTCTGAAG
 451 ATGACARAAG TTTTACTATT CACCAAAGGC AGCTATCTGG TCAACTCTGG
 501 CTTTGACATC GCGAAGGCGA GCGGTCAAGC GCGCAACCTG AGCCGGAGCT
 551 ACGCGATCGT TCGGCGGACC AGCGAACCAG AGGGTCMAGG CTACTTTACC
 601 CACTCTTAGC TCGGCGCTGT TGTTTATACC COTGAAGGCA ACTTCAAAAA
 651 AGTCAGCTTC TCGGACTTGG ACGACGATGC CAANTCCGGN AAATCCGAGG
 701 CCGAATACAT CCGCAAAACC CNGACCGGCT GCGTCGGCAT GATTGAACAC
 751 CACTTCATGT CCACCTGGAT CCTCCAACCC AAGGCGGGAC AAGCGTTTGT
 801 GCGCGCTGGC GACTGCGNGT TNGACATCAA ACGCCGCAAC GACAAGCTGT
 851 ACAGCACCCG CGTCAGCGTG CTTTATAGCG CTATCCAAAA CGGTGCGAAA
 901 TCCNAAGCCT CATCAACCTC CTACGCCGGC CCACAGACCA CATCNGTTAT
 951 CGCAACATC GCGGCAACCC TGCAACTGGN CAAAGACTAC GGCAAGATAC
 1001 ACTGGTTCGC CTCCCCCCTC TTTTGGCTTT TGAACCAACT GCACAACATC
 1051 ATCGGCAACT GGGGCTGGGC GATTATCGTT TTAACCATCA TCGTCAAAAG
 1101 CGTACTGTAT CATTGACCA ACGCCTCTTA CCGTGTAGTG GCGAAAAATG
 1151 GTGCCGCGCG GCCCAAACTG CAAGCCATCA AAGAGAAATA CCGCGACGAC
 1201 CGTATGGCGC AGCAACAACG CATGATCGAT GTTACACACG ACGAGAAATG
 1251 CAACCGCGTG GCGCGCTGCT TCGCTATGCT CTTGCAATCT CCGCTCTTCA
 1301 TCGGCTGCTA TTGCGCGCTG TTGCGCTCGG TAGAATATGC CCACGACRCT
 1351 TSGCTGGGTT GGATTACCGA CCTACGCGCG GCGGACCCCT ACTACATCCT
 1401 GCCCATCATT ATGCGCGCAA CGATGTTCGC CCAAACCTAT CTGAACCCGC
 1451 CGCCGACCGA CCGGATGCAG GCGAAAAATG TGAACATCAT GCCTTTGGTT
 1501 NNTNCNNNNA NGTTCCTCNN CTTCCCTGCC GGCTCTGGAT TGTAAGGGGT
 1551 GATCAACAC CTCTGACCA TCGCCACAGA ATGCGCACATC AACCOCGACGA
 1601 TCGAAAAACA ACGCGCGCAA GCGGAAGTCG TTTCTTAA

This encodes a protein having amino acid sequence <SEQ ID 54>:

1 XDFKRLTXFF AIALVIMIGX XMFPTPKPV PAPOQTAQQQ AVXSAREAL
 51 APXPPIITVTT DTQVAVIDEK SGLRLRLTLL KYKATGDKNK PFILFGDGKX
 101 YTYXAXSELL DAQGNNILKG IGFSAFKQY SLEGDKEVVR LSAPETRGLK
 151 IDKVYTFYTK SYLVNVRFDI ANGSGQTLAN SADYRVVRDH SEPEGQGYFT
 201 HSYVGVVVYV PEGNFQKVSF SLDLDDAXSG KSEAEYIRKT XTWGLMGIEH
 251 HFMTSWILQP KGSQSVCAAG DCXDXIKRRN DKLYSTSVSV PLAAIQNGAK
 301 SXASINLYAG PQTTTSVIANI ALNLQLXKDY GRHWFASSEL FWLIMQLHNI
 351 IGNWGAITV LTIIVKAVLY PLTNASRYSM AKMDAPAEKL QAIKEKYGD
 401 RMAQQAMMO NYTDEKINPL GGCPLPYLQI PVFGLYVAL FASVELRQAP
 451 WLGWITDLRS ADPYIYLPIL MAATMFAQYI LNPFFDPMQ AKMMKIMPLV
 501 SXXXFFXFPA GLVLVWVINN LLTIAQQWHI NRSIEKQRAQ GEVVS*

ORF11a and ORF11-1 show 95.2% identity in 544 aa overlap:

55 orf11a.pep 10 20 30 40 50 60
 XDFKRLTXFFAIALVIMIGXMXMFPTPKVPVPAPOQTAQQQAVXSAREAALAPXXPIITVTT
 ||||| ||||| ||||| ||||| ||||| |||||
 orf11-1 MDFKRLTAFFAIALVIMIGWEKMFPTPKVPVPAQQAQQQAVTASAEAAALAPATPIITVTT
 10 20 30 40 50 60
 60 orf11a.pep 70 80 90 100 110 120
 DTQVAVIDEKSGDLRLRLTLLKYKATGDGNKPFILFGDGKXYTYXAXSELLDAQGNNILKG
 ||||| ||||| ||||| ||||| ||||| |||||
 orf11-1 DTQVAVIDEKSGDLRLRLTLLKYKATGDENKPFILFGDGKEYTYVQASSELLDAQGNNILKG

		70	80	90	100	110	120
		130	140	150	160	170	180
5	orf11a.pep	IGFSAPKKQYSLEGDKVEVRLSAPETRGLKIDKVVYTFTKGSYLVNVRFDIANGSGGTANL					
	orf11-1	IGFSAPKKQYSLEGDKVEVRLSAPETRGLKIDKVVYTFTKGSYLVNVRFDIANGSGGTANL					
		130	140	150	160	170	180
		190	200	210	220	230	240
10	orf11a.pep	SADYRIVRDHSEPEGQGYFTHSYVGPVVYTPEGNFQKVSFSDLDDAKSGKSEAEYIRKT					
	orf11-1	SADYRIVRDHSEPEGQGYFTHSYVGPVVYTPEGNFQKVSFSDLDDAKSGKSEAEYIRKT					
		190	200	210	220	230	240
		250	260	270	280	290	300
15	orf11a.pep	XTGWLGMIEHHFMSTWILQPKGGQSVCAAGDCXXDIKRRNDKLYSTS SVSVP LAAIQNGAK					
	orf11-1	PTGWLGMIEHHFMSTWILQPKGRQSVCAAGECNIDIKRRNDKLYSTS SVSVP LAAIQNGAK					
		250	260	270	280	290	300
20		310	320	330	340	350	360
	orf11a.pep	SXASINLYAGPQTTSVIANIADNLQKDYGKVVWFASPLFWLLNQLHNIIGNWGWAIIV					
	orf11-1	AEASINLYAGPQTTSVIANIADNLQKDYGKVVWFASPLFWLLNQLHNIIGNWGWAIIV					
25		310	320	330	340	350	360
		370	380	390	400	410	420
	orf11a.pep	LTIIVKAVLYPLTNASYRSMAKMRRAAPKLQAIKEKYGDDRMAQQQAMQLYTDEKINFL					
30	orf11-1	LTIIVKAVLYPLTNASYRSMAKMRRAAPKLQAIKEKYGDDRMAQQQAMQLYTDEKINFL					
		370	380	390	400	410	420
		430	440	450	460	470	480
35	orf11a.pep	GGCLMMLLIQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPPYIILPIMAATMFAQTY					
	orf11-1	GGCLMMLLIQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPPYIILPIMAATMFAQTY					
		430	440	450	460	470	480
		490	500	510	520	530	540
40	orf11a.pep	LNPPPTDPMQAKMKIMPLVXSXFFFPAGLVLYWVNNLLTIAQQWHINRSIEKQRAQ					
	orf11-1	LNPPPTDPMQAKMKIMPLVFSVMFFFPAGLVLYWVNNLLTIAQQWHINRSIEKQRAQ					
		490	500	510	520	530	540
45	orf11a.pep	GEVVSX					
	orf11-1	GEVVSX					

Homology with a predicted ORF from *N.gonorrhoeae*

ORF11 shows 96.3% identity over a 240aa overlap with a predicted ORF (ORF11.ng) from *N.gonorrhoeae*:

55	Orf11	NLYAGPQTTSVIANIADNLQKDYGKVVWFASPLFWLLNQLHNIIGNWGWAIIVLT	57
	orf11ng	MAVNLYAGPQTTSVIANIADNLQKDYGKVVWFASPLFWLLNQLHNIIGNWGWAIIVLT	60
	orf11	IIVKAVLYPLTNASYRSMAKMRRAAPKLQAIKEKYGDDRMAQQQAMQLYTDEKINPLGG	117
60	orf11ng	IIVKAVLYPLTNASYRSMAKMRRAAPKLQAIKEKYGDDRMAQQQAMQLYTDEKINPLGG	120
	orf11	CLPMLLIQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPPYIILPIMAATMFAQTYLN	177
65	orf11ng	CLPMLLIQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPPYIILPIMAATMFAQTYLN	180

orf11	PPPTDPMQAKMMKIMPLVFSXKFFFFPAGXVLYWVNNLLTIAQQWHINRSIEKQRAQGE	237
orf11ng	PPPTDPMQAKMMKIMPLVFSVMFFFPAGLVLYWVNNLLTIAQQWHINRSIEKQRAQGE	240
5	orf11	VVS 240
	orf11ng	VVS 243

An ORF11ng nucleotide sequence <SEQ ID 55> was predicted to encode a protein having amino acid sequence <SEQ ID 56>:

10	1	MAVNLYAGPQ	TTSVIANIAD	NLQIAKDYKG	VHWFASPLFW	LLNQLHNIIG
	51	NWGWAIVVLT	IIVKAVLYPL	TNASYRSMAK	MRAAAPELQT	IKEYGDDRM
	101	AQQAMMQLF	EDEEINPLGG	CLPMLLIQIPV	FILGYWALFA	SVELRQAPWL
	151	GWITDLRAD	PYYILPIIMA	ATMFAQTYLN	PPPTDPMQAK	MMKIMPLVFS
	201	VMFFFFPAGL	VLYWVNNLL	TIAQQWHINR	SIEKQRAQGE	VVS*

15 Further sequence analysis revealed the complete gonococcal DNA sequence <SEQ ID 57> to be:

1	ATGGATTTTA	AAAGACTCAC	GGCGTTTTTC	GCCATCGCGC	TGGTGATTAT
51	GATCGGCTGG	GAAAAAATGT	TCCCCACCCG	GAAACCGGTC	CCCGCGCCCC
101	AACAGCGCGC	ACAAAAACAG	GCAGCAACCG	CTTCCGCGCA	AGCCGCGCTC
151	GGCGCCGCAA	GGCGGATTAC	CGTAACGACG	GACACGGTTC	ARGCGGTAT
201	TGCGTGAANA	AGTGGCGAACC	CGGTCGGGCT	GACCGCTGCT	AAATACAAG
251	CAACCGCGCA	CGCAACACAA	CCGCTCGTCC	TGTTTGGCGA	CGGCACAGAA
301	TACACCTTAC	TCGCGCATC	CGAACTTTTG	GACCGCGAGG	GCACACAACT
351	TCGTGAAGGC	ATCGGCTTTA	CGCGACCGAA	AAAACAGTAG	ACCCCTCAACG
401	CGGACACAGT	CGAAGTCCCG	CTGAGCGCGC	CGCAACACAA	CGGACTGAAA
451	ATCGACANAAG	TCTATACCTT	TACCAAGACG	AGCTATCTGG	TCAACGTCGG
501	CTTCGACATC	GCCACGCGCA	CGGCTCAAC	CGCCAACCTG	AGCGCGGACT
551	ACCGCATCGT	CGCGGACAC	AGCGAACCCG	AGGGTCAAGG	CTACTTTACC
601	CACCTCTACG	TCGGCCCTGT	TGTTTATACC	CCTGAAGGCA	ACTTCCAAAA
651	AGTCAGCTTC	TCCGactTtg	acgACGATGC	gaaTCcggc	aaATccgagg
701	cggaatacat	CGCAAAACC	cgacccggtt	ggctcgcgat	gattgaaac
751	cacttcaatg	caacctggat	ctctcAAcct	aaagcgccgc	aaaacgtttg
801	cgccccaggga	gactgcgcta	tcgacattaa	aGcgccgaac	gacaagctgt
851	acagcgcaag	cgtcagcgtg	cctttaaccg	ctatcccacg	ccggggggcca
901	aaacgcaaaa	tggcggtCAA	CCTGTATGCC	GGTCGCAAA	CCACATCGGT
951	TATCGAAAC	ATCGCgga	ACCTGCAACT	GGCAAAAGAC	TACGGTAAAG
1001	TACACTGCTT	CGCATCGCGC	CTCTCTGGC	DKLYSAAACA	ATCTGCACAC
1051	ATTATCGGCA	ACTGGGGCTG	GGCAATCGTC	GTTTTGACCA	TCATCGTCAA
1101	AGCCGTACTG	TATCCATTGA	CCAAACgcctc	ctACCGTTTC	ATGGCGAAAA
1151	TGCGTGccgc	cgcaacCaaA	CTGCAGACCA	TCAAAGAAAA	ATAcgCGAC
1201	GACCGTATGG	CGCAACAGCA	AGCGATGATG	CAGCTTTTCA	AAgacgAGAA
1251	AATCAACCCG	CTGGCGCGCT	GTctgcatat	gctgttgCAA	ATCCCCGTCT
1301	TCATCGGCTT	GTACTGGGCA	TTGTTGCGCT	CGGTAGAATT	GGCGCGAGCA
1351	CCTTGGCTGG	GCTGGATTAC	CGACCTCAGC	CGCGCGGACC	CCTACTACAT
1401	CTCGCCCATC	ATTATGGCGG	CAACGATGTT	CGCCCAACCC	TATCTGAACC
1451	CGCGCGCGAC	CGACCCGATG	CAGGCGAAAA	TGATGAAAA	CATCGCGTGT
1501	GTTTTCTCGG	TCATGTTCTT	CTTCTTCCCT	GCCGGTTTGG	TTTCTCTACTG
1551	GGTGGTCAAC	AACCTCCTGA	CCATCGGCCA	GCAGTGGCAC	ATCAACCGCA
1601	GCATCGAAAA	ACAACGCGCC	CAAGCGAAG	TGTTTTCTCA	A

This encodes a protein having amino acid sequence <SEQ ID 58; ORF11ng-1>:

50	1	MDFKRLTAFF	ALALVIMIGW	EKMFPFPKPV	PAPQQAQKQ	AATASAEAL
	51	APATPTITVT	DTVQAVIDEK	SGDLRLRLTL	KKYKATDENK	PFVLFGDGKE
	101	YTVYQSELL	DAQNNILKG	IGFSAPKKQY	TINGDTVEVR	LSAPEINLKG
	151	IDKVYTFTKD	SYLVNVRFDI	ANGSGQTANL	SADYRIVRDH	SEBEGQGYFT
	201	HSYVGPVVYT	PEGNFQKVSF	SDLDDDAKSG	KSEAEYIRKT	PTGNLGMIEH
55	251	HFMSTWILQP	KGGGNVCAQG	DCRIDIKRRN	DLYSASVSU	PLTAIPTRGP
	301	KPKMAVNLYA	GPQTTSVIAN	IADNLQIAK	YGKVVHFWAS	LFWLLNQLHN
	351	IIGNWGWAIV	VLTIIVKAVL	YPLTNASYRS	MAKMRAPAPK	LQTIKEYKGD
	401	DRMAQQQAMM	QLYKDEKINP	LGCGCLPMLLI	IPFVILGYWA	LFASVELRQA
	451	PWLGWITDLS	RADFPYIILPI	IMAAATMFAQT	YLNPPPTDPM	QAKMMKIMPL
60	501	VFSVMFFFFP	AGLVLYWVNN	NLLTIAQQWH	INRSIEKQRA	QGEVVS*

ORF11ng-1 and ORF11-1 shown 95.1% identity in 546 aa overlap:

		10	20	30	40	50	60
	orf11ng-1.pep	MDFKRLTAFFAIALVIMIGWEKMFPTFKVPAPQQAQKAATASAAALAPATPITVTT					
5	orf11-1	MDFKRLTAFFAIALVIMIGWEKMFPTFKVPAPQQAQKAATASAAALAPATPITVTT					
		10	20	30	40	50	60
	orf11ng-1.pep	DTVQVAIDEKSGDLRLRLTLKYKATGDENKPFVLFGDGKEYTYVAQSELLDAQNNILKG					
10	orf11-1	DTVQVAIDEKSGDLRLRLTLKYKATGDENKPFVLFGDGKEYTYVAQSELLDAQNNILKG					
		70	80	90	100	110	120
	orf11ng-1.pep	IGFSAPKKQYTLNGDITVEVRLSAPETNGLKIDKVVYFTKDSYLVNVRFDIANGSGQTANL					
15	orf11-1	IGFSAPKKQYTLNGDITVEVRLSAPETNGLKIDKVVYFTKDSYLVNVRFDIANGSGQTANL					
		130	140	150	160	170	180
	orf11ng-1.pep	SADYRIVRDHSEPEGQGYFTHSYVGPVVYTPEGNFQKVSFSLDDDDAKSGKSEAEYIRKT					
20	orf11-1	SADYRIVRDHSEPEGQGYFTHSYVGPVVYTPEGNFQKVSFSLDDDDAKSGKSEAEYIRKT					
		190	200	210	220	230	240
	orf11ng-1.pep	PTGWLGMIEHHFMSTWILQPKGGQNVCAQGDRCRIDIKRRNDKLYSASVSPLEITRGP					
25	orf11-1	PTGWLGMIEHHFMSTWILQPKGGQNVCAQGDRCRIDIKRRNDKLYSASVSPLEITRGP					
		250	260	270	280	290	300
	orf11ng-1.pep	KPKMAVNLVYAGFQTTSVIANIADNLQAKDYKGVHWFASPLFWLLNLQHLNIGNWGWAIV					
30	orf11-1	KPKMAVNLVYAGFQTTSVIANIADNLQAKDYKGVHWFASPLFWLLNLQHLNIGNWGWAIV					
		310	320	330	340	350	360
	orf11ng-1.pep	VLTIIVKAVLYPLTNASYRSMKMRAPKLTQIKEKYGDDRMQAQQAMMOLYKDEKINP					
35	orf11-1	VLTIIVKAVLYPLTNASYRSMKMRAPKLTQIKEKYGDDRMQAQQAMMOLYKDEKINP					
		370	380	390	400	410	420
	orf11ng-1.pep	LGGCLPMLLQIPVFIGLYWALFASVELRQAFWLGWITDLSRADFPYILPLIIMAAITMAFQT					
40	orf11-1	LGGCLPMLLQIPVFIGLYWALFASVELRQAFWLGWITDLSRADFPYILPLIIMAAITMAFQT					
		430	440	450	460	470	480
	orf11ng-1.pep	YLNPPPTDPMQAKMMKIMELVFSVMFFFFFAGLVLYVNVNLLTIAQQWHINRSIEKQRA					
45	orf11-1	YLNPPPTDPMQAKMMKIMELVFSVMFFFFFAGLVLYVNVNLLTIAQQWHINRSIEKQRA					
		490	500	510	520	530	540
	orf11ng-1.pep	QGEVVVSX					
50	orf11-1	QGEVVVSX					
		540					

60 In addition, ORF11ng-1 shows significant homology with an inner-membrane protein from the database (accession number p25754):

65 ID 60IM PSEFU STANDARD; PRT: 560 AA.
AC F25754;
DT 01-MAY-1992 (REL. 22, CREATED)
DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE 60 KD INNER-MEMBRANE PROTEIN. . . .

SCORES Initl: 1074 Initn: 1293 Opt: 1103
 Smith-Waterman score: 1406; 41.5% identity in 574 aa overlap

```

5      10      20      30      40
orf11ng-1.pep MDFKR---LTAFFAIALVIMIGW-----EKMFFT-----PKVPFAPQQAQKQ
      ||:||  ::|: :: |::| : ||  | ||| ::|: :
p25754 MDIKRTILIAALAVVSVMVLKWNDDYGOAALPTQNTAASTVAEGLPDGVPGAGNNGASAD
      10      20      30      40      50      60

10     50     60     70     80     90
orf11ng-1.pep AATASAEALAPATPIT-----VTTDTVQAVIDEKSGDLRLRLLLKYKATGDE-NKPF
      :|:|::| :|::| :|::| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|
p25754 VPSANAESEPAELAPVALSKDLIRVKTDLVLELAIDFVGGDIVQLNLKPYPRRQGHNPPIF
      70     80     90     100    110    120

15     100    110    120    130    140
orf11ng-1.pep VLFPGKEITYVAQSELLDAQGNILKIG---FSAPKKQYTL-NGD---TVEVRLSAFE
      ||:|| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|
p25754 QLFNDNGGERVYLAQSGLTGTDGPDRA-RASGRFLYAECKSYQLADGQQLVVDLKFSS---
      130    140    150    160    170

20     150    160    170    180    190    200
orf11ng-1.pep TNGLIKDKVYTFTKDSYLVNVRFDIANGSGQTANLSADYRIVRDHS-EPEGQGYF-THSY
      ||::| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|
p25754 DNGVNYIKRFSFKRGEYDLNVSYLIDQSGQAWNGMFAQLKRDASGDPSSTATGTATY
      180    190    200    210    220    230

25     210    220    230    240    250    260
orf11ng-1.pep VGFVVYTPPEGNFQKVSFSDLDLDDAKSGKSEAEYIRKTPWTGLMIEHHFMSTWLQPKGG
      :|:|::| :|:|:|:|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|
p25754 LGAALWTASEPYKKVMKIDID--KGSLEK----NVSGGVAVWVQLQHYFVTAWI-PAKSD
      240    250    260    270    280

30     270    280    290    300    310    320
orf11ng-1.pep QNVCAQGDCCRDIKRRNDKLYSASVSPVLTAPTRGPKPKMAVNLYAGPQTTSVIANIAD
      :|| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|
p25754 NNV-----VQTRKDSQGYIIGYTGFWISVPA-GGKVETSALLYAGPKIQSKLKLSP
      290    300    310    320    330

35     330    340    350    360    370    380
orf11ng-1.pep NLQLAKDYGKVHWF-ASPLFWLLNQLHNIIGNWGWAIIVVLTIVKAVLYPLTNASYRMA
      :|:|:| :|:| :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
p25754 GLELTVDYGFL-WFIAQPIFWLLQHINSLLGNWSIIVLTMLIKGLGFLFSAASYRMA
      340    350    360    370    380    390

40     390    400    410    420    430    440
orf11ng-1.pep KMPAAAPKQTIKEKYGDDRMAQQQAMQLYKDEKINPLGGCLPLMLQIPVFIPLYWALF
      :|:|:|:|:| :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
p25754 RMRVAPKLAALKERFGDDRKMSQAMMELYKKKEKINPLGGCLPLVQMVEFLAYWILL
      400    410    420    430    440    450

45     450    460    470    480    490    500
orf11ng-1.pep ASVELRQAPFLGWITDLRADPYYILPIIIMAAFTMAQTYLNPPTDPMQAKMMKIMPLVF
      :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
p25754 ESVEMRQAPWILWITDLSIKDPFFILPIIMGATMFIIQQRLNPTPPDPMQAKYMMKMPIIF
      460    470    480    490    500    510

50     510    520    530    540
orf11ng-1.pep SVMFFFPFAGLVLYVWVNNLLTIAQQWHINRSIEKQRAQGEVVSX
      :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
p25754 TFFFLWFFPAGLVLYVWVNNCLSSQQWYITRRIEAATKAAA
      520    530    540    550    560

```

Based on this analysis, including the homology to an inner-membrane protein from *P. putida* and the predicted transmembrane domains (seen in both the meningococcal and gonococcal proteins), it

is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 8

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 59>:

```

5      1  ..GCCGCTCTTAA TCATCGAATT ATTGACGGGA ACGGTTTATC TTTTGGTTGT
      51  NAGCGCGGCT TTGGCGGGTT CGGCGATTGC TTACGGGCTG ACCGGCAGTA
     101  CGCCTGCCGC CGTCTTGACC GNCGCTCTGC TTTCCGGCTG GGGTATTING
     151  TTGCTACACG CCAAAACCGC CGTTAGAAAA GTTGAAACGG ATTCATATCA
     201  GGATTTCGAT GCCGACAAT ATGTCGAAAT CCTCCGNCAC ACAGGCGGCA
     10  251  ACCGTACGTA AGTT.TTTAT CGCGGTACG. ACTGGCAGGC TCAAAATAGC
     301  GGGCAGAGAAG AGCTTGAAAC AGGAACCTGC GCCCTCATTG TCCGCAGGA
     351  AGGCAACCTT CTTATTATCA CACACCCTTA A
  
```

This corresponds to the amino acid sequence <SEQ ID 60; ORF13>:

```

15      1  ..AVLIIELLTG TVYLLVVSAA LAGSGIAYGL TGSTPAAVLT XALLSALGIX
      51  FVHAKTAVRK VETDSYQDLG AGQYVEILRH TGGNRYEVY RGTWQAQNT
     101  GQEELEPGTR ALIVRKEGNL LIITHP*
  
```

Further sequence analysis elaborated the DNA sequence slightly <SEQ ID 61>:

```

20      1  ..GCCGCTCTTAA TCATCGAATT ATTGACGGGA ACGGTTTATC TTTTGGTTGT
      51  NAGCGCGGCT TTGGCGGGTT CGGCGATTGC TTACGGGCTG ACCGGCAGTA
     101  CGCCTGCCGC CGTCTTGACC GNCGCTCTGC TTTCCGGCTG GGGTATTING
     151  TTGCTACACG CCAAAACCGC CGTTAGAAAA GTTGAAACGG ATTCATATCA
     201  GGATTTCGAT GCCGACAAT ATGTCGAAAT CCTCCGNCAC ACAGGCGGCA
     251  ACCGTACGTA AGTTTCTTAT CGCGGTACG. ACTGGCAGGC TCAAAATAGC
     301  GGGCAGAGAAG AGCTTGAAAC AGGAACCTGC GCCCTCATTG TCCGCAGGA
     351  AGGCAACCTT CTTATTATCA CACACCCTTA A
  
```

This corresponds to the amino acid sequence <SEQ ID 62; ORF13-1>:

```

30      1  ..AVLIIELLTG TVYLLVVSAA LAGSGIAYGL TGSTPAAVLT XALLSALGIX
      51  FVHAKTAVRK VETDSYQDLG AGQYVEILRH TGGNRYEVY RGTWQAQNT
     101  GQEELEPGTR ALIVRKEGNL LIITHP*
  
```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF13 shows 92.9% identity over a 126aa overlap with an ORF (ORF13a) from strain A of *N.*

meningitidis:

```

35      orf13.pep      10      20      30      40      50
      orf13a      AVLIIELLTGTVYLLVVSAA LAGSGIAYGLTGSTPAAVLT XALLSALGIXF
      MTWVFWVA AVAVLIIELLTGTVYLLVVSAA LAGSGIAYGLTGSTPAAVLT XALLSALGIXF
      10      20      30      40      50      60
40      orf13.pep      60      70      80      90      100      110
      orf13a      VHAKTAVRKVETDSYQDLGAGQYVEILRH TGGNRYEVY RGTWQAQNT GQEELEPGTRA
      VHAKTAVRKVETDSYQDLGAGQYVEILRH TGGNRYEVY RGTWQAQNT GQEELEPGTRA
      70      80      90      100      110      120
45      orf13.pep      120
      orf13a      LIVRKEGNLLIITHPX
      LIVRKEGNLLIIAKPK
50      130
  
```

The complete length ORF13a nucleotide sequence <SEQ ID 63> is:

```

1  ATGACTGTAT  GGTTCGTGTC  CGCTGTTGCC  GTCTTAATCA  TCGAATTATT
51  GACGGGAACG  GTTATCTTTT  TGGTTGTCAG  CGCGGCTTTG  GCGGTTTCGG
101 GCATTGCTTAA  CGGGCTGACC  GGCAGCACGC  CTGCGCCGCT  CTTGACCGCC
151 GCTCTGCTTT  CCGGCTGGG  TATTGTTTC  GTACACGCCA  AAACCGCGCT
201 GGGAAAAGTT  GAAACGGATT  CATATCAGGA  TTTGGATGCC  GGGCAATATG
251 CGAAATCCT  CCGGCACGCA  GCGGCGAACC  GTTACGAAGT  TTTTATCGC
301 GGTACGCACT  GGCAGGCTCA  AATACGGGG  CAAGAAGAGC  TTGAACGAG
351 AACGCGCGCC  CTAATCGTCC  GCAAGGAAGG  CAACCTTCCT  ATCATGCAA
401 AACCTTAA

```

This encodes a protein having amino acid sequence <SEQ ID 64>:

```

1  MTVWFVAAVA  VLIIELLTGT  VYLLVVSAL  AGSGIAYGLT  GSTPAAVLTA
51  ALLSALGIWF  VHAKTAVKGV  ETDSYQDLDA  QYAEILRHA  GGNRYEVFYR
101 GTHWQAQNTG  QEELFPGTRA  LIVRKEGNLL  IIAKP*

```

ORF13a and ORF13-1 show 94.4% identity in 126 aa overlap

```

10      20      30      40      50      60
orf13a.pep  MTVWFVAAVAVLIIELLTGTVYLLVVSALAGSGIAYGLTGSTPAAVLTAALLSALGIWF
10      20      30      40      50
orf13-1      AVLIIELLTGTVYLLVVSALAGSGIAYGLTGSTPAAVLTXALLSALGIXF

70      80      90      100     110     120
orf13a.pep  VHAKTAVRKVETDSYQDLDAQYAEILRHTGGNRYEVFYRGTHWQAQNTGQEELFPGTRA
25      30      40      50      60      70      80      90      100     110
orf13-1      VHAKTAVRKVETDSYQDLDAQYAEILRHTGGNRYEVFYRGTHWQAQNTGQEELFPGTRA
60      70      80      90      100     110

130
orf13a.pep  LIVRKEGNLLIIAKPX
30      35      40      45      50
orf13-1      LIVRKEGNLLIITHPX
120

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF13 shows 89.7% identity over a 126aa overlap with a predicted ORF (ORF13.ng) from *N. gonorrhoeae*:

```

orf13      AVLIIELLTGTVYLLVVSALAGSGIAYGLTGSTPAAVLTXALLSALGIXF  51
40  orf13ng  MTVWFVAAVAVLIIELLTGTVYLLVVSALAGSGIAYGLTGSTPAAVLTAALLSALGIWF  60

orf13      VHAKTAVRKVETDSYQDLDAQYAEILRHTGGNRYEVFYRGTHWQAQNTGQEELFPGTRA  111
45  orf13ng  VHAKTAVRKVETDSYQDLDAQYAEILRHTGGNRYEVFYRGTHWQAQNTGQEVFEPGTRA  120

orf13      LIVRKEGNLLIITHP  126
45  orf13ng  LIVRKEGNLLIITNP  135

```

The complete length ORF13ng nucleotide sequence <SEQ ID 65> is:

```

1  ATGACTGTAT  GGTTCGTGTC  CGCTGTTGCC  GTCTTAATCA  TCGAATTATT
50  51  GACGGGAACG  GTTATCTTTT  TGGTTGTCAG  CGCGGCTTTG  GCGGTTTCGG
101 GCATTGCTTAA  CGGGCTGACC  GGCAGCACGC  CTGCGCCGCT  CTTGACCGCC
151 GCACTGCTTT  CCGGCTGGG  TATTGTTTC  GTACATGCCA  AAACCGCGCT
201 GGGAAAAGTT  GAAACGGATT  CATATCAGGA  TTTGGATACC  GGAAAATATG
251 CGAAATCCT  CCGATACACA  GCGGCGAACC  GTTACGAAGT  TTTTATCGC
55  301 GGTACGCACT  GGCAGGCGCA  AATACGGGG  CAGGAAGTGT  TTGAACGCGG
351 AACGCGCGCC  CTCATCGTCC  GCAAGAAGG  TAACCTTCCT  ATCATCGCAA
401 ACCCTTAA

```

This encodes a protein having amino acid sequence <SEQ ID 66>:

```

1  MTWVFVAAVA VLIIELLTGT VYLLVVSAAAL AGSGIAYGLT GSTPAAVLTA
51  ALLSALGIWF VHAKTAVGKV ETDSYQDLDT GKYAEILRYT GGNRYEVFYR
101 GTHWQAQNTG QEVFEPGTRA LIVRKEGNLL IIANP*

```

5 ORF13ng shows 91.3% identity in 126 aa overlap with ORF13-1:

```

                                10      20      30      40      50
orf13-1.pep      AVLIEELLTGTVYLLVVSAAALAGSGIAYGLTGSTPAAVLTKALLSALGIXF
                                |||||
orf13ng      MTWVFVAARVAVLIEELLTGTVYLLVVSAAALAGSGIAYGLTGSTPAAVLTKALLSALGIWF
10
                                60      70      80      90      100     110
orf13-1.pep      VHAKTAVRKVETDSYQDLDAQYVEILRHTGGNRYEVFYRGTHWQAQNTGQEELEPGTRA
15  orf13ng      VHAKTAVGKVETDSYQDLDTGKYAEILRYTGGNRYEVFYRGTHWQAQNTGQEVFEPGTRA
                                |||||
                                70      80      90      100     110     120
orf13-1.pep      LIVRKEGNLLIITHPX
20  orf13ng      LIVRKEGNLLIANPX
                                |||||
                                120
                                130

```

Based on this analysis, including the extensive leader sequence in this protein, it is predicted that ORF13 and ORF13ng are likely to be outer membrane proteins. It is thus predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 9

The following DNA sequence was identified in *N.meningitidis* <SEQ ID 67>:

```

30 1  ATGTWTGATT TCGGTTTtGG CGArCTGGTT TTTGTGCGCA TTATCGCCCT
51 GATWtGCCTC GGCcCCGAAC GCsTGCCCGA GGCCGCCCGC AycCGCGGAC
101 GGcTCATCGG CAGGCTGCAA GCCTTtGTcG GcAGCGTCAA ACAGGAATTT
151 GACACTCAAA TCGAACTGGA AGAACTGAGG AAGGCAAAAC AGGAATTTGA
35 201 AGCTGCCGcC GCTCAGGTTC GAGACAGCCT CAAGAARACC GTTACGGATA
251 TGGAAAGGCAA TCTGCAcGAC ATTTCCGACG GTCTGAAGCC TTGGGAAAAA
301 CTGCCCGAaC AGCGGACACC TGCCGATTTC GGTGTGcGATG AAAACGGCAA
351 TCCGCT.TCC CGATGCGGCA AACACCCCTAT CAGACGGCAT TTCCGAGCTT
401 ATGCGCTC..

```

This corresponds to the amino acid sequence <SEQ ID 68; ORF2>:

```

40 1  MXDFGLGELV FVGIIALIVL GPERXPAAAR XAGRLIGRLQ RFVGSVKQEF
51 DTQIELEELR KAKQEFEEAA AQVRDSLKET GTDMbGNLHD ISDGLKPWEK
101 LPEQRTPADF GVDENGNPXS RCGKHPfIRRH FRRYAV..

```

Further work revealed the complete nucleotide sequence <SEQ ID 69>:

```

45 1  ATGTTTGATT TCGGTTTGGG CGAGCTGGTT TTTGTGCGCA TTATCGCCCT
51 GATTGTCCCTC GGCcCCGAAC GCCTGCCCGA GGCCGCCCGC ACCGCGGAC
101 GGCTCATCGG CAGGCTGCAA GCCTTtGTcG GCAGCGTCAA ACAGGAATTT
151 GACACTCAAA TCGAACTGGA AGAACTGAGG AAGGCAAAAC AGGAATTTGA
201 AGCTGCCGCC GCTCAGGTTC GAGACAGCCT CAAGAARACC GTTACGGATA
50 251 TGGAAAGGCAA TCTGCAcGAC ATTTCCGACG GTCTGAAGCC TTGGGAAAAA
301 CTGCCCGAaC AGCGGACACC TGCCGATTTC GGTGTGcGATG AAAACGGCAA
351 TCCGCTTCCG GAYGCGCAA ACACCCCTATC AGACGGCATT TCCGACGTTA
401 TGCCGTCGCA ACGTTCCTAC GCTTCCGCGC AARCCCTTGG GGACAGCGGG

```

5
10
451 CAAACCGGCA GTACAGCCGA ACCCGCGGAA ACGACCAAG ACGCGCATG
501 GCGGGAATAC CTGACTGCTT CTGCCGCGC ACCCGTCGTA CAGACCGTGG
551 AAGTCAGCTA TATCGATACT GCTGTTGAAA CGCGTGTTC GCACACCACT
601 TCCTGCGCA AACAGGCAAT AAGCGCGAAA CGCGATTTC GTCCGAAACA
651 CCGCGCGAAA CCTAAATTGC GCGTCCGTAA ATCAATA

This corresponds to the amino acid sequence <SEQ ID 70; ORF2-1>:

1 MFDGFLGELV FVGIIALIVL GPERLPEAAR TAGRLIGRLQ RFVGSVKQEF
51 DTQIELEELR KAKQEFEEAA AQVRDSLKET GTDMEGNLHD ISDGLKPWEK
101 LPEQRTPADF GVDENGNPLP DAANTLSDGI SDVMPERSY ASAETLGDGSG
151 QTGSTAEPAE TDQDRAWREY LTASAAAPVV QTVEVSYIDT AVETVPVPHHT
201 SLRKQAISRK RDLRPKHKRAK PKLRVRKS*

Further work identified the corresponding gene in strain A of *N.meningitidis* <SEQ ID 71 >:

1 ATGTTTGATT TCGGTTTGGG CGAGCTGGTT TTTGTCGGCA TTATCGCCCT
51 GATTGTCCTC GGCCCGGAAC GCCTGCGCGA GCGCGCCGCG ACCGCGCGGAC
101 GGCTCATCGG CAGGCTGCAA CGCTTTGTGC GCAGCGTCAA ACAGGAATTT
151 GACACGCAAA TCGAACTGGA AGRACCTAAG AAGCGAAAGC AGGAATTTGA
201 AGCTGCGGCT GCTCAGGTTT GAGACAGCCT CAAGAAGAAC GGTACGGATA
251 TGGAGGGTAA TCTGCACGAC ATTTCCGACG GTCCTGAAGC TTGGGAAAAA
301 CTGCGCGAAC AGCGCGCGCC TCGTGTATTC GGTGTCGATG AAACACGCGAA
351 TCCTGTTTCC GATCGCGCAA ACACCGCTAT AGACGCGATC TCCGACGTTA
401 TGCGGTCGGA ACCTGCTTAC GCTTCCGCGG AAGCGCTTGG GAGACGCGGG
451 CAAACCGGCA GTACAGCCGA ACCCGCGGAA ACCGACCAAG ACCGTGCGATG
501 GCGGGAATAC CTGACTGCTT CTGCCGCGCG ACCCGTCGTA CAGACCGTGG
551 AAGTCAGCTA TATCGATACT CTGTTTGAAT CCGCTGTTCC GCATACCACT
601 TCCTGCGCTA AACAGGCAAT AAGCGCGAAA CGCGATTTC GTCTAAATC
651 CCGCGCGAAA CCTAAATTGC GCGTCCGTAA ATCAATA

This encodes a protein having amino acid sequence <SEQ ID 72; ORF2a>:

1 MFDGFLGELV FVGIIALIVL GPERLPEAAR TAGRLIGRLQ RFVGSVKQEF
51 DTQIELEELR KAKQEFEEAA AQVRDSLKET GTDMEGNLHD ISDGLKPWEK
101 LPEQRTPADF GVDENGNPFP DAANTLLDGI SDVMPERSY ASAETLGDGSG
151 QTGSTAEPAE TDQDRAWREY LTASAAAPVV QTVEVSYIDT AVETVPVPHHT
201 SLRKQAISRK RDLRPKSRK PKLRVRKS*

The originally-identified partial strain B sequence (ORF2) shows 97.5% identity over a 118aa overlap with ORF2a:

35
40
45
50
orf2.pep MDGFLGELVFVGIIALIVLGPXPERKPEAARXAGRLIGRLQRFVGSVKQEFDTQIELEELR
orf2a MFDGFLGELVFVGIIALIVLGPRLPEAARTAGRLIGRLQRFVGSVKQEFDTQIELEELR
orf2.pep KAKQEFEEAAQAQRVDSLKETGTDMENLHDISDGLKPWEKLEPQRTPADFGVDENGNPXS
orf2a KAKQEFEEAAQAQRVDSLKETGTDMENLHDISDGLKPWEKLEPQRTPADFGVDENGNPFF
orf2.pep RCGKHPIRRHFRRYAV
orf2a DAANTLLDGISDVMPERSYASAE TLGDSGGTGSTAEPAETDQDRAWREYLTASAAAPVV

The complete strain B sequence (ORF2-1) and ORF2a show 98.2% identity in 228 aa overlap:

55
orf2a.pep MFDGFLGELVFVGIIALIVLGPRLPEAARTAGRLIGRLQRFVGSVKQEFDTQIELEELR 60
orf2-1 MFDGFLGELVFVGIIALIVLGPRLPEAARTAGRLIGRLQRFVGSVKQEFDTQIELEELR 60
orf2a.pep KAKQEFEEAAQAQRVDSLKETGTDMENLHDISDGLKPWEKLEPQRTPADFGVDENGNPFF 120

```

orf2-1      KAKQEFAAAAQVRDLSKETGTDMGNI.LHIISDGLKFWKLPQRTPADPGVDENGNIPLF 120
orf2a.pep   DAANTILLDGISVMPFSERSYASAEITLGDSCGTGTAETPAETQDRAWREYLTASAAAPVV 180
orf2-1      |||||
orf2-1      DANTILSDGISVMPFSERSYASAEITLGDSCGTGTAETPAETQDRAWREYLTASAAAPVV 180
orf2a.pep   |||||
orf2a.pep   QTVEVSYIDTAVETVPVPHHTLSLKKQAISRRKDLRPKRAKPKLVRVKSX 229
orf2-1      |||||
orf2-1      QTVEVSYIDTAVETVPVPHHTLSLKKQAISRRKDLRPKRAKPKLVRVKSX 229

```

Further work identified a partial DNA sequence <SEQ ID 73> in *N.gonorrhoeae* encoding the following amino acid sequence <SEQ ID 74; ORF2ng>:

1 MFDFGLGELI FVGIIALIVL GPERLPEAAR TAGRLIGRLQ RFVGSVKQEL
51 DTQIELEELR KVKQAFESAA AQVRDSLKET DTDQNSLHD ISDGLKPWEK
101 LPEORTPADF GVDEKGNLSL RYGKHRRRH FRRYAV*

Further work identified the complete gonococcal gene sequence <SEO ID 75>:

1	ATGTTGTTGATT	CGGTTTGGG	CGAGCTGATT	TGTTCCGG	TTATCGGAT
5	GATGTTCGTT	GCTGCAGAAC	GCTGCCCGA	AGGCGCCGC	ACTCGCGCAC
10	GGCTTATCGG	CAGGACTGCA	AGCTTTGTAG	GAAGCTCAAG	CAAGAAACAT
15	GACACTCAAA	CGACTGTGA	AGAGCTGAGG	AAGGCTCA	AGGCATTCGA
20	AGCTGCGCCG	GCTCAGGCT	GAGACAGGCT	CAAGAACCC	GATACAGGATA
25	TCGAGAACAG	TCTGCAGAC	ATTTCGAGC	GCTCAAGCC	TGGGAAATTA
30	CTGCGCGAAG	AGGCGACGC	tcgcgatttc	gATGCGATT	AAACcgggaa
35	tcctctctcc	gATCGGATA	ACACCGTATC	AGTCCGCA	TCGACAGTAA
40	TCGCGCTCGA	AGCTTCGAT	ACTTccgCG	AAACCTGAT	GGAGCAGGAT
45	CAAACCGCA	GTACAGCGA	CTGCGCGAA	AGCCACAAG	ACCGCGCATG
50	CGGGAATAC	CTAGctctct	ctgcgcgcgc	acctgctgta	Cagagggcgc
55	tcgaagatct	ctataTCGAT	ATGCTGTGTG	AAAGcctgct	tcgacacCac
60	actcttcctg	gcAACACGCG	AATPAAACCG	TAACCGGATT	Tttgtccgaa
65	ACACCGCGCG	aAACCGGAA	tcgcctctCG	TAAATCCTAA	

This encodes a protein having the amino acid sequence <SEO ID 76: ORF2ng-1>:

1	MFDFGLGEL	FVGIIALIVL	GPERLPEAAR	TAGRLIGRLQ	RFVGSVKQEL
51	DQIQIEELR	KVQAEFAAA	AQRVDSIKET	SDVMNSLHD	ISDLGLPWEK
101	LEPQRTPADF	GVDGNPLP	DTANTVSDGI	TDMPNSERSD	TSAAETLGDGR
151	QTGSKAPEAE	TKDKRAWREY	KLAKAAAPVV	QRAVEVSYID	TAVETVPVHT
201	LRGKTOAIRN	KRDECPKHRA	PTPRLVRSKS		

The originally-identified partial strain B sequence (ORF2) shows 87.5% identity over a 136aa overlap with ORF2ng:

orf2.pep	MXDFGLGELVFGVIGIATLVLPGRXPFAARXAGRLIGRLQRFVGSVQKQDFDTQIELEELR	60
orf2.ng	MFDFGLGELVFGVIGIATLVLPGRPLPFAARTAGRLIGRLQRFVGSVQKQLDTQIELEELR	60
orf2.pep	KAKEFEAAAAQVRDSLKETGDTMEGNLHDSIDGLKPWEKLPEQRTPADFGVDENGNXFS	120
orf2.ng	KVKQFEAAAAQVRDSLKETDTMDQNSLHDSIDGLKPWEKLPEQRTPADFGVDENGNLSP	120
orf2.pep	RCGKHPIRRHFRRYAV	136
orf2.ng	RYGKHPIRRHFRRYAV	136

The complete strain B and gonococcal sequences (ORF2-1 & ORF2ng-1) show 91.7% identity in 229 aa overlap:

orf2-1.pep MFD FGLGELVFVGI IALIV LGPERLPEAARTAGRLIGRLQRFVGSVKQEDTQIELEELR
 orf2ng-1 MFD FGLGELVFVGI IALIV LGPERLPEAARTAGRLIGRLQRFVGSVKQEDTQIELEELR

		10	20	30	40	50	60
5	orf2-1-pep	70	80	90	100	110	120
		KAKQEFEEAAAQVRDSLKETGTDMGNLHDISDGLKPWEKLPEQRTPADFGVDENGNPLP					
	orf2ng-1	70	80	90	100	110	120
		KVKQAFEAAAQVRDSLKETDTDMQNSLHDISDGLKPWEKLPEQRTPADFGVDENGNPLP					
10	orf2-1-pep	130	140	150	160	170	180
		DAANTLSDGISDVMPSESRYSASAE TLGDSGGTGSTAEP AE TDQD RAWREYLTASAAAPV					
	orf2ng-1	130	140	150	160	170	180
		DTANTVSDGISDVMPSESRSDTSAE TLGDDRTGSTAEP AE TDK RAWREYLTASAAAPV					
15	orf2-1-pep	190	200	210	220	229	
		Q-TVEVSYIDTAVETVPVPH TTSLRKQAI SRKRDFR PKHRAKPKLR VK KSX					
	orf2ng-1	190	200	210	220	230	
		QRAVEVSYIDTAVETVPVPH TTSLRKQAI NRKRDFC PKHRAKPKLR VK KSX					

- 20 Computer analysis of these amino acid sequences indicates a transmembrane region (underlined), and also revealed homology (59% identity) between the gonococcal sequence and the TatB protein of *E.coli*:

25 gnl|PID|e1292181 (AJ005830) TatB protein [Escherichia coli] Length = 171
 Score = 56.6 bits (134), Expect = 1e-07
 Identities = 30/88 (34%), Positives = 52/88 (59%), Gaps = 1/88 (1%)

30 Query: 1 MFDPLGELIFVGIIALIVLGPRLPEAARTAGRLIGRLQRFVGSVKQELDTQIELEELR 60
 MFD G EL+ V II L+VLGP+RLP A +T I L+ +V+ EL ++L+E+ +
 Sbjct: 1 MFDIGFSELLVFTIIGLVVLGPQRLVFAVKTAGVIRALRSLATTVQNELTQELKIQEFQ 60

Query: 61 -KVKQAFEAAAQVRDSLKETDTDMQNS 87
 +K+ +A+ + LK + +++ +
 Sbjct: 61 DSLKRVKASLTNLTPELKASMDLRQA 88

- 35 Based on this analysis, it was predicted that ORF2, ORF2a and ORF2ng are likely to be membrane proteins and so the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

- ORF2-1 (16kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 3A shows the results of affinity purification of the GST-fusion protein, and Figure 3B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for Western blots (Figure 3C), ELISA (positive result), and FACS analysis (Figure 3D). These experiments confirm that ORF37-1 is a surface-exposed protein, and that it is a useful immunogen.

Example 10

- 45 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 77>:

1 ATGCAAGCAC GGCTGCTGAT ACCTATTCTT TTTTCAGTTT TTTATTTATC
 51 CGC.TGCGGG ACACTGACAG GTATTCCATC GCATGGCGgA GkTAAACgCt

5
 10
 101 TTGCGGTGCA ACAAGAACTT GTGGCGCTT CTGCCAGAGC TGCCGTTAAA
 151 GACATGGATT TACAGGCATT ACACGGACGA AAGATTGCAT TGTACATTGC
 201 CACTATGGGC GACCAAGGTT CAGGCAAGTT GACAGGGGGG TCGCTACTCC
 251 ATTGATGCAC KGTTCGCTGG CGAATACATA ACAGCCCTCG CCGTCCGTAG
 301 CGATTACACT TATCCACGTT ACGAACCAC GCTCGAACA ACATCGAAGC
 351 GTTTGACAGG TTTAACCACT TCTTTATCTA CACTTAAATGC CCCTGCACCT
 401 TCTCGCACCC AATCAGACGG TAGCGAAGT AAAAGCAGTC TGGGCTTAAA
 451 TATTGGCGGG ATGGGGGATT ATCGAAATGA AACCTGACG ACTAACCOCG
 501 GCGACACTGC CTTTCTTTCC CACTTGGTAC AGACCGTATT TTTCCTGGCG
 551 GGCATAGAGC TTGTTTCTCC TGCCAATGCC GATACAGATG TGTTTATTAA
 601 CATCGACGTA TTGGAACGA TAGCAACAG AACCGAAATG..

This corresponds to the amino acid sequence <SEQ ID 78; ORF15>:

15
 1 MQARLLIPIL FSVFILSACG TLTGIPSHGG XKRFAVEQEL VAASARAARK
 51 DMDLQALHGR KVALYIATMG DQSGSGLTGG RYSIDAXXGG EYINSPAVRT
 101 DYTYPREYET AETTSGLLTG LTLSLSTLNA PALSRITQSDG SGSKSSLGLN
 151 IGGMGDYRNE TLTTNPRDTA FLSHLVQTFV FLRGIDVVPV ANADTDVFIN
 201 IDVFGTIRNR TEM..

Further work revealed the complete nucleotide sequence <SEQ ID 79>:

20
 25
 30
 35
 40
 45
 1 ATGCAAGCAC GGCTGCTGAT ACCTATTCTT TTTTCAGTTT TTAATTTATC
 51 CGCGTGCGGG ACACGACGAG GTATTCCATC GCATGGCGGA GGTAAACGCT
 101 TTGCGGTGCA ACAAGAACTT GTGGCGCTT CTGCCAGAGC TGCCGTTAAA
 151 GACATGGATT TACAGGCATT ACACGGACGA AAGATTGCAT TGTACATTGC
 201 CACTATGGGC GACCAAGGTT CAGGCAAGTT GACAGGGGGT CGCTACTCCA
 251 TTGATGCACT GATTTCGTGGC GAATACATAA ACAGCCCTCG CCGTCCGTAG
 301 GATTACACCT ATCCACGTTA CGAAACCACC GCTGAAACAA CATCAGGGGG
 351 TTTGACAGGT TTAACCACTT CTTTATCTAC ACTTAATGCC CTTGCACTCT
 401 CTGCAACCCA ATCAGACGGT AGCGGAAGTA AAAGCAGTCT GGGCTTAAAT
 451 ATTGGCGGGA TGGGGGATTA TCGAAATGAA ACCTTGACGA CTAACCCGGG
 501 CGACACTGCC TTTCTTTCCC ACTTGGTACA GACCGTATT TTCTCGCGCG
 551 GCATAGACGT TGTTCCTCCT GCCAATTGCG ATACAGATGT GTTATTATAC
 601 ATCCAGCGAT TCGGAACGAT ACGCAACAGA ACCGAAATGC ACCTATACAA
 651 TGCCGAACCA CTGAAGAGCC AAACAAACTT GGAATATTTC CAGATAGACA
 701 GACCAATATA AAAATTGCTC ATCAACACAA AACCAATATG GTTTGAAGCT
 751 GCCTATAAAG AAAATTAGCG ATTGTGGATG GGGCGGTATN AATATAGCAA
 801 AGGAATTAAA CGACGCGAAG GATTAAATGT CCAATTTCTC GATATCCGAC
 851 CATACGGCAA TCATACGGGT AACTCCGCC CACTCGTAGA GGCTGATAAC
 901 AGTCATGAGG GGTATGGATA CAGCGATGAA GTAGTGCAC AACATAGACA
 951 AGGACAACCT TGA

This corresponds to the amino acid sequence <SEQ ID 80; ORF15-1>:

40
 45
 1 MQARLLIPIL FSVFILSACG TLTGIPSHGG GKRFVEQEL VAASARAARK
 51 DMDLQALHGR KVALYIATMG DQSGSGLTGG RYSIDALIRG EYINSPAVRT
 101 DYTYPREYET AETTSGLLTG LTLSLSTLNA PALSRITQSDG SGSKSSLGLN
 151 IGGMGDYRNE TLTTNPRDTA FLSHLVQTFV FLRGIDVVPV ANADTDVFIN
 201 IDVFGTIRNR TEMHLNAPT LKAQTKLEYF AVORTNKKLL IKPKTNAFEA
 251 AYKENYALWM GPYKVKSGIK FTEGLMVDFF DIRPIGNHTG NSAPVEADN
 301 SHEGYGYSDE VVRQHRQGP *

Further work identified the corresponding gene in strain A of *N.meningitidis* <SEQ ID 81>:

50
 55
 60
 1 ATGCAAGCAC GGCTGCTGAT ACCTATTCTT TTTTCAGTTT TTAATTTATC
 51 CGCGTGCGGG ACACGACGAG GTATTCCATC GCATGGCGGA GGTAAACGCT
 101 TTGCGGTGCA ACAAGAACTT GTGGCGCTT CTGCCAGAGC TGCCGTTAAA
 151 GACATGGATT TACAGGCATT ACACGGACGA AAGATTGCAT TGTACATTGC
 201 AACTATGGGC GACCAAGGTT CAGGCAAGTT GACAGGGGGT CGCTACTCCA
 251 TTGATGCACT GATTTCGTGGC GAATACATAA ACAGCCCTCG CCGTCCGTAG
 301 GATTACACCT ATCCACGTTA CGAAACCACC CTTTATCTAC ACTTAATGCC CTTGCACTCT
 351 TTTGACAGGT ATCCACGTTA AGCGGAAGTA AAAGCAGTCT GGGCTTAAAT
 401 CGCGCAACCA ATCCACGTTA TCGAAATGAA ACCTTGACGA CTAACCCGGG
 451 ATTGGCGGGA TGGGGGATTA TCGAAATGAA ACCTTGACGA CTAACCCGGG
 501 CGACACTGCC TTTCTTTCCC ACTTGGTACA GACCGTATT TTCTCGCGCG
 551 GCATAGACGT TGTTCCTCCT GCCAATTGCG ATACAGATGT GTTATTATAC
 601 ATCGACGAT TCGGAACGAT ACGCAACAGA ACCGAAATGC ACCTATACAA
 651 TGCCGAACCA CTGAAGAGCC AAACAAACTT GGAATATTTC CAGATAGACA

5 701 GAACCAATAA AAAATTGCTC ATCAAACCAA AAACCAATGC GTTGAAGCT
 751 GCCTATAAAG AAAATTACGC ATTGTGGATT GGACCGTATA AAGTAAGCAA
 801 AGGAATTAA CCGACACAGG GATTAAATGCT CGATTTCGCC GATATCCAC
 851 CATACGCGAA TCATATGSGT AACTCTGCCC CATCCGTAGA GGCTGATAAC
 901 AGTCATGAGG GGTATGGATA CAGCGATGAA CGATGCGAC GACATAGACA
 951 AGGCAACCT TGA

This encodes a protein having amino acid sequence <SEQ ID 82; ORF15a>:

10 1 MQARLLIPLIL FSVFILSACG TLTGIPSHGG KRFSAVEQEL VAASARAAVK
 51 DMDLQALHGR KVALYIATMG DQSGSGSLTGG RYSIDALIRG EYINSPAVRT
 101 DYTYPYRSETT AETTSGGTLG LTSSLSTLINA PALSRQTQSDG SGSKSSSLGLN
 151 IGGMGDYRNE TLTTNPRDTA FLSHLVQTVF FLRGIDVVSP ANADTDVFIN
 201 IDVFGTIRNR TEMHLYNAET LKAQTKLEYF AVDRTNKKLL IKPKTNAFEA
 251 AYKENYALWM GPYKVSQKIG PTEGLMVDVS DIQPYGNHMG NSAPSVVEADN
 301 SHEGYGYSDE AVRHRHQGP *

15 The originally-identified partial strain B sequence (ORF15) shows 98.1% identity over a 213aa overlap with ORF15a:

		10	20	30	40	50	60
orf15.pep		MQARLLIPLILFSVFILSACGTLTGIPSHGGKRFSAVEQELVAASARAAVKMDLQALHGR					
orf15a		MQARLLIPLILFSVFILSACGTLTGIPSHGGKRFSAVEQELVAASARAAVKMDLQALHGR					
		10	20	30	40	50	60
orf15.pep		70	80	90	100	110	120
orf15a		KVALYIATMGDQSGSGSLTGGRYSIDALIRGEYINSPAVRTDYTPYRSETTAEETSSGLTG					
		70	80	90	100	110	120
orf15.pep		130	140	150	160	170	180
orf15a		LTSSLSTLNAPALSRQTQSDGSGSKSSSLGLNIGMGDYRNETLTNPRDTAFLSHLVQTVF					
		130	140	150	160	170	180
orf15.pep		190	200	210			
orf15a		FLRGIDVVSPANADTDVFINIDVFGTIRNRTEM					
		190	200	210	220	230	240
orf15a		FLRGIDVVSPANADTDVFINIDVFGTIRNRTEMHLYNAETLKAQTKLEYFAVDRTNKKLL					

40 The complete strain B sequence (ORF15-1) and ORF15a show 98.8% identity in 320 aa overlap:

		10	20	30	40	50	60
orf15a.pep		MQARLLIPLILFSVFILSACGTLTGIPSHGGKRFSAVEQELVAASARAAVKMDLQALHGR					
orf15-1		MQARLLIPLILFSVFILSACGTLTGIPSHGGKRFSAVEQELVAASARAAVKMDLQALHGR					
		10	20	30	40	50	60
orf15a.pep		70	80	90	100	110	120
orf15-1		KVALYIATMGDQSGSGSLTGGRYSIDALIRGEYINSPAVRTDYTPYRSETTAEETSSGLTG					
		70	80	90	100	110	120
orf15a.pep		130	140	150	160	170	180
orf15-1		LTSSLSTLNAPALSRQTQSDGSGSKSSSLGLNIGMGDYRNETLTNPRDTAFLSHLVQTVF					
		130	140	150	160	170	180
orf15a.pep		190	200	210	220	230	240
orf15-1		FLRGIDVVSPANADTDVFINIDVFGTIRNRTEMHLYNAETLKAQTKLEYFAVDRTNKKLL					

		190	200	210	220	230	240
		250	260	270	280	290	300
5	orf15a.pep	IKPKTNAFEAAKYKENYALWMGPYKVSKGIKPTEGLMVD FSDIQPYGNHMGNSAPSV EADN					
	orf15-1	IKPKTNAFEAAKYKENYALWMGPYKVSKGIKPTEGLMVD FSDIRPYGNHMGNSAPSV EADN					
		250	260	270	280	290	300
10	orf15a.pep	SHEGYGYSDEAVRRHRQQQFX					
	orf15-1	SHEGYGYSDEVVRQHRQQQFX					
		310	320				
		310	320				

Further work identified the corresponding gene in *N.gonorrhoeae* <SEQ ID 83>:

15	1	ATGCGGGCAC	GGCTGCTGAT	ACCTATTCTT	TTTTCAGTTT	TTATTTTATC
	51	CGCCTGGGGG	ACACTGACAG	GTATTCGATC	GCATGGCGGA	GGCAACGCT
	101	TGCGGSGTGA	ACACAGCACT	CTGGCCGCTT	CTGCCAGAGC	TGCCGTAAAT
	151	GACATGATT	TACAGGCATT	ACACGGACGA	AAAGTTGCAT	TGTACATTGC
	201	AACATATGSC	GACCAAGGTT	CAGGCAGTTT	GACAGGGGGT	CGCTACTCCA
20	251	TTGATGCACT	GATTGCGCGG	GAATACATAA	ACAGCCCTCG	CGTCGCGACC
	301	GATTACACCT	ATCCGCGTTA	CGAACCACAC	GCTGAAACAA	CATCAGGGCG
	351	TTTACGCGGT	TTAACCACCT	CTTTATCTAT	ACTTAATGCC	CTGCATCTCT
	401	CGCGCACCCA	ATCAGACGGT	AGCGGAAGTA	GGAGCAGTCT	GGGCTTAAT
	451	ATTGGCGGGA	TGGGGGATTA	TCGAATGAA	ACCTTGACGA	CCAACCCGGG
25	501	CGACACTGCC	TTTCTTTCCC	ACTTGGTGCA	GACCGTATTT	TTCTCGCGCG
	551	GCATAGACGT	TGTTTCTCCT	GCCAATGCCG	ATACAGATGT	GTTTATTATC
	601	ATCGACGTAT	TCGGAACGAT	ACGCAACAGA	ACCGAAATGC	ACCTATACAA
	651	TCCCGAAACA	CTGAAGCCCG	AAACAAACT	GGAATATTTC	GCACTAGACA
	701	GAACCAATAA	AAATTTGCTC	ATCAAAACCA	AAACCAATGC	GTTTGAAGCT
30	751	GCCTATAAAG	AAAATTACGC	ATTGTGGATG	GGGCCGTATA	AAGTAGACAA
	801	AGGAATCAAA	CCGACGGGAG	GATTGATGGT	CGATTTCTCC	GATATCCAC
	851	CATACGGCAA	TCATACGGGT	AACCTCCGCC	CATCCGTAGA	GCTGATATC
	901	AGTCATGAGG	GGTATGGATA	CAGCGATGAA	SCAGTCGCAC	AACATAGACA
	951	AGGGCAACCT	TGA			

This encodes a protein having amino acid sequence <SEQ ID 84; ORF15ng>:

35	1	MRARLLIPIL	FSVFILSACG	TLTGIPSHGG	GKRFVEQEL	VAASARAARK
	51	DMDLQALHGR	KVALYIATMG	DQSGSLTGG	RYSIDALIRG	EYINSPAVRT
	101	DYTYPRYETT	AETTSGGLTG	LTTSLSTLNA	PALSRTQSDG	SGSRSSSLGLN
	151	IGGMDGYRNE	TLTNPFRDTA	FLSHLVQTVF	FLRGIDVVP	ANADTVFVN
40	201	IDVFGTIRNR	TEMHLYNAET	LKAQTKLEYF	AVDRTNKKLL	IKPKTNAFEA
	251	AYKENYALWM	GPYKVSIGIK	PTEGLMVD	FIQPYGNHGT	NSAPSV EADN
	301	SHEGYGYSDE	AVRQHRQGP	*		

The originally-identified partial strain B sequence (ORF15) shows 97.2% identity over a 213aa overlap with ORF15ng:

45	orf15.pep	MQARLLIPILFSVFILSACGTLTGIPSHGGKRFVEQELVAASARAARKDMDLQALHGR	60
	orf15ng	MRARLLIPILFSVFILSACGTLTGIPSHGGKRFVEQELVAASARAARKDMDLQALHGR	60
50	orf15.pep	KVALYIATMGDQSGSLTGGGRYSIDAXXGEYINSAPVRTDYTYPRYETTAETTSGLT	120
	orf15ng	KVALYIATMGDQSGSLTGGGRYSIDALIRGEYINSAPVRTDYTYPRYETTAETTSGLT	120
	orf15.pep	LTTSLSTLNPALSRQSDGSGSKSLGLNIGMGDYRNETLTTNPRDTAFLSHLVQTVF	180
55	orf15ng	LTTSLSTLNPALSRQSDGSGSRSSSLGLNIGMGDYRNETLTTNPRDTAFLSHLVQTVF	180
	orf15.pep	FLRGIDVVPANADTDVFINIDVFGTIRNRTEM	213
	orf15ng	FLRGIDVVPANADTDVFINIDVFGTIRNRTEMHLYNAETLKAQTKLEYFAVDRTNKKLL	240

The complete strain B sequence (ORF15-1) and ORF15ng show 98.8% identity in 320 aa overlap:

-113-

		10	20	30	40	50	60
	orf15-1.pep	MQARLLIPILPSVFILSAGCTLTGIPSHGGGKRF	AVEQELVAASARA	AVKMDLQALHGR			
5	orf15ng	MRARLLIPILPSVFILSAGCTLTGIPSHGGGKRF	AVEQELVAASARA	AVKMDLQALHGR			
		10	20	30	40	50	60
	orf15-1.pep	KVALYIATMGDQSGSLTGGRYSIDALIRGEYINSP	AVRVDYTPRYETTAETTS	GGTLG			
10	orf15ng	KVALYIATMGDQSGSLTGGRYSIDALIRGEYINSP	AVRVDYTPRYETTAETTS	GGTLG			
		70	80	90	100	110	120
	orf15-1.pep	LTTSLSTLNAPALSRTQSDGSGSKSSLGLNIGMG	DYRNETLTTPRDTAFSL	SHLVQTVF			
15	orf15ng	LTTSLSTLNAPALSRTQSDGSGSKSSLGLNIGMG	DYRNETLTTPRDTAFSL	SHLVQTVF			
		130	140	150	160	170	180
	orf15-1.pep	FLRGIDVVSANADTDVFINIDVFGTIRNRTEMH	LYNAETLKAQTKLEYFA	VDRTNKKLL			
20	orf15ng	FLRGIDVVSANADTDVFINIDVFGTIRNRTEMH	LYNAETLKAQTKLEYFA	VDRTNKKLL			
		190	200	210	220	230	240
	orf15-1.pep	IKPKTNAFEAAAYKENYALWMGPYKVS	KGIKPTEGLMVD	FSDIRPYGNHTGNS	SAPSV	VEADN	
25	orf15ng	IKPKTNAFEAAAYKENYALWMGPYKVS	KGIKPTEGLMVD	FSDIRPYGNHTGNS	SAPSV	VEADN	
		250	260	270	280	290	300
	orf15-1.pep	SHEGYGYSDEVVRQHRQGGQFX					
30	orf15ng	SHEGYGYSDEVVRQHRQGGQFX					
		310	320				
35	orf15-1.pep						
	orf15ng						

Computer analysis of these amino acid sequences reveals an ILSAC motif (putative membrane lipoprotein lipid attachment site, as predicted by the MOTIFS program).

indicates a putative leader sequence, and it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF15-1 (31.7kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 4A shows the results of affinity purification of the GST-fusion protein, and Figure 4B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for Western blot (Figure 4C) and ELISA (positive result). These experiments confirm that ORFX-1 is a surface-exposed protein, and that it is a useful immunogen.

Example 11

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 85>:

```

1  ..GG.CAGCACAAAAACAGGC GGTGAACGG AAAAACGTA TTACGATGA
51  TGCGGGGATAT GATATTCGCG GTATTCACGG GCGATTCTC CGCAAAATAT

```

101 ATCCCCGGGT TCGGGCTTCA AATTTCTTC ATCCTGTTTT TAACCGCGGT
 151 GGCATTCAAA ACACCTGCATA CGACCCCTCA GACGGCATCC CGCCCGCTGC
 201 CCGGACTGCC CCGACTGACT GCGGTTTCCA CACTGTTCGG CACAAATGTCG
 251 CTGCGGCTGC GCATAGCGGG CGGTTCACCT TCGTCCCTCT TCTTAATCCA
 301 CTGCGGCTTC CCGCCGCATA AAGCAATCGG CACCATCCCG GGCCCTTGCTT
 351 GGCGGATGCG ACTCTCCGGC GCAATATCGT ATCTGCTCAA CGGCTGTAAT
 401 ATTGCAAGAT TGCCCGAAGG GTCACCTGGG TCCCTTTACC TGCCCGCGGT
 451 CGCCGCTCCT AGCGCGGCAA CCAATTGCCT TGCCCGCGTC GGTGTCAAAA
 501 CCGCCACAAA ACTTCTCTCT GCCAACTCA AAAAATC.TT CGGCATTATG
 10 551 TTGCTTTTGA TTGCGGAAA AATGCTGTAC AACTCGCTTT AA

This corresponds to the amino acid sequence <SEQ ID 86; ORF17>:

1 ..GQHKKQAVNG KTVFTMPGMI IFGVFTGAFS AKYIPAFGLQ IFFILFLTA
 51 AFKTLHTDPQ TASRPLGLPL XLTAVSTLFG TMSSWVGIGG GSLSVPLFIH
 101 CGFPAHKAIG TSSGLAWPIA LSGAISYLLN GLNIAGLPEG SLGLFYLPVAV
 151 AVLSAATIAF AFLGVKTAHK LSSAKLKSF GIMLLIAGK MLNLL*

Further work revealed the complete nucleotide sequence <SEQ ID 87>:

1 ATGTGGCATT GGGACATTAT CTTAATCCTG CTGCGCTAGC GCAGTGCGGC
 51 AGGTTTATT GCGCGCTGT TCGGCTAGG CGCGCGCAOG CTGATTGTCC
 101 CTGCTGTTTT ATGGGTCGCT GATTTCGAGG GTTTGGCACA ACATCCTTAC
 151 GCGCAACACG TCGCACTCTC CCGCTCATCG TCTTCACCGC TCTTCACCGC
 201 CTTTCCAGAT ATGCTGGGCG AGCACAAAAA ACAGCGCGTC GACTCGAA
 251 CCGTATTACG GATGATGCGG GGTATGATAT TCGGCTATT CACGGCGGCA
 301 CTCTCCGCAA AATATATCCC CGCGTTCGGG CTCAAAATTT TCTTCATCTC
 351 GTTTTAAACC GCGCTCGCAT TCAAAACACT GCATACCGAC CCTCAGACGG
 25 401 CATCCCGCCC GTCGCCGCGA CTGCCCGGAG TGACTCGGTT TTCCACACTG
 451 TTCGGCACAA GTGCGAGCTG GGTGCGGATA GCGCGCGGTT CACTTTCGTT
 501 CCCCTTCTTA ATCCACTGCG GCTTCCCGCG CCATAAAGCC ATCGGCACAT
 551 CATCCGGCCT TGCCCTGGCGG ATTGCACTCT CGGCGCGCAT ATCGTATCTG
 601 CTCACCGCCC TGAATATTGC AGGATTGCCG GAGGGGTAC TGGGCTTCTC
 30 651 TTACCTGCCG CGCGTCGCGG TCCTCAGCGG GGCAACCAT GCGTTTGGCC
 701 CGCTCGGTGT CAAAACGCGC CACAAACTTT CTTCTGCCAA ACTCAAAAAA
 751 Tc.TTCGGCA TTATGTGCTG TTTGATTGCC GGGAAAAATG GTGACAACTC
 801 GCTTTAA

This corresponds to the amino acid sequence <SEQ ID 88; ORF17-1>:

1 MWHWDIILIL LAVGSAAGFI AGLEGVGGGT LIVEVVLWLV DLQGLAQHPY
 51 AQHLAVGTSE AVMVPTAFSS MLGQHKQAV DWKTVFTMP GMIFGVFTGA
 101 LSAKYIPAFG LQIFFILFLT AVAFKTLHTD PQTASRPLFG FLGLTAVSTL
 151 FGTSSWVGII GGGSLSVFLL THCGFPAHKA IGTSSGLAWP IALSGAISYL
 40 201 LNLGLIAGLP EGSLGLFYLP AVAVLSAATI AFAPLGVKTA HKLSSAKLKK
 251 XFGIMLLIA GKMLYNLL*

Computer analysis of this amino acid sequence gave the following results:

Homology with hypothetical *H. influenzae* transmembrane protein HI0902 (accession number P44070)

ORF17 and HI0902 proteins show 28% aa identity in 192 aa overlap:

ORF17 3 HKKQAVNGKTVFTMPGMI IFGVFTGAFS AKYIPAFGLQIF--FILFLTAVAFKTLHTDP 59
 HK + + V + P ++ VF G F + +IF ++L ++ D
 HI0902 72 HKLGIWVQAVRIAPVIMLSVFIQGLFIFGRDLREISAKIFACLVVYLATRMVLISIKKD- 130
 ORF17 60 QTASRPLGLPLXLTAVSTLFGTSSWVGIGGGSLSVFPLHCGFPAHKAIGTSSGLAWPI 119
 Q ++ L L + L G SS GIGGG VPFL G +AIG+S+ +
 HI0902 131 QVTTKSLTPLSSVIG-GILIGMASSAAGIGGGGFIVPFLTARGINIKQAIGSSAFQGMLL 189
 ORF17 120 ALSGAISYLLNGLNIAGLPEGSGLFYLPVAVAVLSAATIAFAPLGVXXXXXXXXXXXXX 179
 +SG S++++G +FE SLG++YLEAV +A + + LG
 HI0902 190 GISGMFSPISVSGWGNPLMEYSIGLYIYLPVAVLGITATSFSTSKLGASATAKLPVSTLKKG 249
 55 ORF17 180 FGIMLLIAGKM 191
 F + L++++ M
 HI0902 250 FALFLIVVAIM 261

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF17 shows 96.9% identity over a 196aa overlap with an ORF (ORF17a) from strain A of *N. meningitidis*:

[illegible]

The complete length ORF17a nucleotide sequence <SEQ ID 89> is:

1	ATGTGGCATT	GGGACATTAT	CTTATCTCTC	CTTGGCGGTAG	CGAGTGGCGGC
51	AGGTTTCTATT	GGGCTCTCTG	TGGCTGCTAG	CGTGGCGACG	CGATTGTCCTG
101	CTGTGCTGTT	ATGCTGCTCTT	GATTCTCAGG	TTTGGGCACA	ACATTCTCTTAC
151	GGCGACACAC	TCCGCTCTCG	GAATCATCTG	CGCGTCATGG	CTTTCACCGG
201	CTTTTCCAGT	ATGCTGGGGC	GGGACAAAAA	ACAGGCGGTC	GACTGGAAAGA
251	CGGATATGAT	GATGATGCGC	GGTATGGTAT	TCGGCGTATT	CGCTGGCCGCA
301	CTCTCCGACA	AATATATCCC	AGCGTTCTGG	CTTCAAAATT	TCCTTCATCTC
351	TTTATTAAAC	GCGCTCGGAT	TCAAACAATC	GCATACCGAC	CTCCAGACGG
401	CATCCGCGCC	GCTGCGGCGA	CTGCCCGGAC	TGACTCGGCT	TTCCACATGT
451	TTCCGACACA	TTGTCGAGTA	GGTCCGCGTA	GGGGCGCGGT	CACTTTCTCGT
501	CCCTCTCTTA	ATCCACTGCT	GTTCCTCCGC	CTCAAAAGCG	ATCCGCGACAT
551	CATCCGGCTC	TGCTGTGGCG	AGGTCATCTC	CCGGCGCGAT	ATCTGATGAT
601	CTCAACGGCC	TGAATATTGC	AGGATTTATC	GAGGGTCAC	TGGCGTCTGC
651	TACTCTGCTC	TCTCTGCTGC	TCTCTGCTGC	TCGCTGCTGC	GCGTTCTGCC
701	GCGTGGGTGT	CAAAACGCC	CAAAACATT	CTTCTGCCAA	ACTCAAAAAA
751	TCCTTGGCGA	TGATGTTGCT	TTTGATTGCG	GGAAAAATTC	TGTACAACTT
801	GCCTTT&&				

45 This encodes a protein having amino acid sequence <SEQ ID 90>:

50

1	MWHWDIILIL	LVGSSAAGFI	AGLFGVGGGT	LIVPVVLWL	DQLGSLVFA
51	AQHLAVGTSE	AVVETAFSS	MQLGHRQAV	DWKTFTMMP	MGVGFVPGA
101	LSAKVPAEGL	LVGSLVFA	AVAEKFLHTD	PQTASRPLPG	LPGLTAVSTL
151	FTGTMSSWGI	CGGSLVPEPL	IHOGFPAHKA	IGTSSGLPWL	LSLGSALVY
201	INGNLINAGL	EGSLGFLYLP	AVAVLSAATI	AFAPLGVKTA	HKLSAKLKK
251	SLNLTMLLTA	GKMLYLLI*			

ORE17a and ORE17-1 show 98.9% identity in 268 aa overlap:

```

55      10      20      30      40      50      60
      orf17a.pep  MWHWDII LILLAVGSAAGFIAGLFGVGGGTLIVPVVLWVLDLQGLAQHPYQAHLAVGTSF
      orf17-1     MWHWDII LILLAVGSAAGFIAGLFGVGGGTLIVPVVLWVLDLQGLAQHPYQAHLAVGTSF
      10      20      30      40      50      60
60      orf17a.pep  AVVMVTAFSMSMLGQHKHQAQDWKTVTMTMPEGMVFGVGAQASAKYIPAGLQIFILELT

```


551 CATCCGGCCT TGCCTGCGCG ATTCGACTCT CCGGGCGAAT ATCGTATCTG
 601 CTCACGGTC TGAATATTGC AGGATTGCCG GAAGGGTCGC TGGGCTTCTC
 651 TTACCTGCCC CGCCTGCGCG TCCTCAGCGC GGCAACCATT GCCTTTGCCG
 701 CGCTCGGTGT CAAACCGCC CACAACTTTT CTTCGCCAA ACTCAAGAA
 751 TCCTTCGGCA TTATGTTGCT TTTGATTGCC GAAAAATGC TGTACAACCT
 801 GCTTTAA

This corresponds to the amino acid sequence <SEQ ID 94; ORF17ng-1>:

10 1 MHHWDIILIL LAVGSAAGFI AGLEFGVGGGT LIVPVVLWL DLQGLAQHPY
 51 AQHLAVGTSE AVMVFTAFSS MLGQHKQAV DWKTIFAMMP GMIFGVFAGA
 101 LSAKYIPAFG LQIFFILFLT AVAFKTLHTG RQTASRPLPG LPGLTAVSTL
 151 FGAMSSWVG I GGGSLVPFL IHCGFPAHKA ITGSSGLAWP IALSGAISYL
 201 VNGLNIAGLP EGSGLFLYLP AVAVLSAATI AFAPLGVKTA HKLSSAKLKE
 251 SFGIMLLLLTA GKMLYNLL*

ORF17ng-1 and ORF17-1 show 96.6% identity in 268 aa overlap:

15 orf17-1.pep 10 20 30 40 50 60
 orf17ng-1 MHHWDIILILAVGSAAGFIAGLEFGVGGGT LIVPVVLWLDLQGLAQHPYQHLAVGTSF
 10 20 30 40 50 60
 70 80 90 100 110 120
 orf17-1.pep AVMVFTAFSSMLGQHKQAVDWKTIVFTMMPGMIFGVFTGALSAKYIPAFGLQIFFILFLT
 orf17ng-1 AVMVFTAFSSMLGQHKQAVDWKTIFAMMPGMIFGVFAGALSAKYIPAFGLQIFFILFLT
 25 70 80 90 100 110 120
 130 140 150 160 170 180
 orf17-1.pep AVAFKTLHTDPQTASRPLPGLTAVSTLFGTMSSWVGIGGSSLPVFLIHCFFPAHKA
 orf17ng-1 AVAFKTLHTGRQTASRPLPGLTAVSTLFGAMSSWVGIGGSSLPVFLIHCFFPAHKA
 30 130 140 150 160 170 180
 190 200 210 220 230 240
 orf17-1.pep IGTSSGLAWPIALSGAISYLLNGLNIAGLEPGSLGFLYLPAAVAVLSAATIAFAPLGVKTA
 orf17ng-1 IGTSSGLAWPIALSGAISYLVNGLNIAGLEPGSLGFLYLPAAVAVLSAATIAFAPLGVKTA
 35 190 200 210 220 230 240
 250 260 269
 40 orf17-1.pep HKLSSAKLKKXFGIMLLLIAGKMLYNLLX
 orf17ng-1 HKLSSAKLKESEFGIMLLLIAGKMLYNLLX
 250 260

In addition, ORF17ng-1 shows significant homology with a hypothetical *H. influenzae* protein:

45 sp|P44070|Y902_HAEIN HYPOTHETICAL PROTEIN HI0902 pir|G64015 hypothetical protein
 HI0902 - Haemophilus influenzae (strain Rd KW20) gi|1573922 (U32772) H. influenzae
 predicted coding region HI0902 [Haemophilus influenzae] length = 264
 Score = 74 (34.9 bits), Expect = 1.6e-23, Sum P(2) = 1.6e-23
 Identities = 15/43 (34%), Positives = 23/43 (53%)
 50 Query: 55 AVGTSFAVMVFTAFSSMLGQHKQAVDWKTIFAMMPGMIFGVF 97
 A+GTSFA +V T S HK + W+ + + P++ VF
 Sbjct: 52 ALGTSFATIVITGIGSAQRHHKLGNIWQAVRILAPVMSLVE 94
 55 Score = 195 (91.9 bits), Expect = 1.6e-23, Sum P(2) = 1.6e-23
 Identities = 44/114 (38%), Positives = 65/114 (57%)
 60 Query: 150 LFGAMSSWVGIGGSSLPVFLIHCFFPAHKAIGTSSGLAWPIALSGAISYLVNGLNIAGL 209
 L G SS GIGG VPFL G +AIG+ + +SG S++V+ +
 Sbjct: 148 LIGMASSAAGIGGGFIVPFLTAGRINIKQAIGSSAFCGMLLIGSGMFSFVSGWGNPL 207
 Query: 210 PEGSLGFLYLPAAVAVLSAATIAFAPLGVKTAHKLSSAKLKESEFGIMLLLIAGKM 263
 PE SLG++YLPVAV ++A + + LG KL + LK+ F + L+++ A M

Subjct: 208 PEYSLGYIYLPVAVLGITATSFFTSKLGASATAKLPVSTLKKGFALFIVVAINM 261

This analysis, including the homology with the hypothetical *H.influenzae* transmembrane protein, suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 12

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 95>:

```

1  ..GGAAACGGAT  GGCAGGCAGA  CCCCAGACAT  CCGCTGCTCG  GGCTTTTTGC
51  CGTCAGTAAT  GTATCGATGA  CGCTTGCTTT  TGTCGGAATA  TGTGCGTTGG
101  TGCATTATTG  CTTTTCGGGA  ACGGTTCAAG  TGTTTGTGTT  TCGGGCACTG
151  CTCAAACTTT  ATGCGCTGAA  GCGCGTTTAT  TGGTTCGTGT  TGCAGTTTGT
201  GCTGATGCGC  GTTGCCATAT  TCCACCGCTG  CGGTATAGAC  CGGCAGCCGC
251  CGTCAACGTT  CGGCGGCTCG  CAGCTCGGAC  CTTCGCGGTT  GACGCGAGCG
301  TTGATGCAGG  TCTCGGTACT  GGTGCTGCTG  CTTTCAAGAA  TTGAAGATA
351  A

```

This corresponds to the amino acid sequence <SEQ ID 96; ORF18>:

```

1  ..GNGWQADPEH  PLLGLEAVSN  VSMPLAFVGI  CALVHYCFSG  TVQVFVFAAL
51  LKLYALKFVY  WFLVQFVIMA  VAYVHRCGID  RQPPSTFGGS  QLRLGSLTAA
101  LMQVSVLVLL  LSEIGR*

```

Further work revealed the complete nucleotide sequence <SEQ ID 97>:

```

1  ATGATTTTGC  TGCATTGGA  TTTTGTGCT  GCCTTACTGT  ATGCGGCGGT
51  TTTTCTGTTT  CTGATATCC  GCGCAGGAAT  GTTGCAATGG  TTTTGGGCGA
101  GTATTATGCT  GTGGCTGGGC  ATATCGGTTT  TGGGGGCAAA  GCTGATGCC
151  GGCATATGGG  GAATGACCCG  GCGCGCGCCC  TTGTTCATCC  CCCATTTTAA
201  CCGTACTTTG  GGCAGCATAT  TTTTTTTCAT  CGGGCATTGG  AACCGGAAAA
251  CAGATGGAAA  CGGATGGCAG  GCAGACCCCG  AACATCCGCT  GCTCGGCGTT
301  TTTGCCGTCA  GTAATGTATC  GATGACGCTT  GCTTTTGTGC  GAATATGTGC
351  GTTGGTGCA  TATTGCTTTT  CGGGAACGGT  TCAAGTGTTT  GTGTTTGCAG
401  CACTGCTCAA  ACTTTATGCG  CTGAAGCCGG  TTTATTTGGT  CGTGTTCGAG
451  TTTGTGCTGA  TGGCGGTGTC  CTATGTCAC  CGCTCGGGTA  TAGACCGSCA
501  GCGCGCGTCA  ACGTTCGGCG  GCTCGCAGCT  GCGACTCGGC  GGGTTGACGG
551  CAGCGTTGAT  GCAGGTCTCG  GTACTGGTGC  TGCTGCTTTC  AGAATTGGA
601  AGATAA

```

This corresponds to the amino acid sequence <SEQ ID 98; ORF18-1>:

```

1  MILLHLDLFL  ALIYAAYFLF  LIFRAGMLQW  FWASIMLNLG  ISVLGAKIMP
51  GIWGMTRAAP  LEIPHLYTL  GSIFFFIHGW  NRKTDGNGWQ  ADPEHPLLGL
101  FAYSNVSMTL  AFVGICALVH  YCFSGTVQVF  VFAALLKLYA  LKPVYWFVLQ
151  FVLMAVAVYH  RCGIDRQPPS  TFGGSQRLRG  GLTALAMQVS  VLVLLEIQL
201  R*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF18 shows 98.3% identity over a 116aa overlap with an ORF (ORF18a) from strain A of *N.meningitidis*:

```

45  orf18.pep                                10      20      30
orf18a  TRAAPLFIPHLYTLGSIFFFIHWNKRTDNGWQADPEHPLLGLFAVSNVSMPLAFVGI
        60      70      80      90     100     110

```


		40	50	60	70	80	90
orf18.pep		CALVHYCFSGTVQVVFVAALLKLYALKPVYWFVLQFVLMAYVYVHRCGIDRQPPSTFGGS					
5	orf18a	CALVHYCFSGTVQVVFVAALLKLYALKPVYWFVLQFVLMAYVYVHRCGIDRQPPSTFGGS					
		120	130	140	150	160	170
		100	110				
10	orf18.pep	QLRLGGLTAALMQSVLVLLVLLSEIGRX					
	orf18a	QLRLGGLTAALMQSVLVLLVLLSEIGRX					
		180	190	200			

The complete length ORF18a nucleotide sequence <SEQ ID 99> is:

	1	ATGATTTTGC	TGCATTGGA	TTTTTTGTCT	GCCTTACTGT	ATCGCGCGGT
15	51	TTTTCTGTTT	CTGATATTCC	GCGCAGGAAT	GTGCAATGG	TTTTGGCGGA
	101	GTATTATGCT	GTGGCTGGCC	ATATCGGTTT	TGGGGSCAAA	GCTGATGCCC
	151	GGCATATGGG	GAATGACCCG	CGCCGCGCCC	TGTTGATCC	CCCATTTTAA
	201	CCGACATTTG	GGCAGCATAT	TTTTTTTCAT	CGGCAATGG	AACCGGAAAA
	251	CGGATGGAAA	CGGATGGCAG	CGAGACCCCG	AACATCCCT	GCTCGGGCTG
20	301	TTTGCCTGCA	GTAATGTATC	GATGACGCTT	GCTTTTGTGG	GAATATGTGG
	351	GTTGGTGCAT	TATTGCTTTT	CGNAGCCGT	TCAAGTGTTT	GTGTTTGGCG
	401	CACCTGCTCA	ACTTTATGCG	CTGAAGCCGG	TTTATTGGTT	CGTGTTCGAG
	451	TTTGTGCTGA	TGGCGGTGTC	CTATGTCCAC	CGCTGCGGTA	TAGACCGGCA
	501	GCGCCCTGCA	ACGTTCCGCG	GNTCGCAGCT	CGGACTCGGC	GGGTTGACGG
25	551	CAGCGTTGAT	GCAGNTCTCG	GTACTGTGTC	TGCTGCTTTC	AGAAATTTGA
	601	AGATAA				

This encodes a protein having amino acid sequence <SEQ ID 100>:

	1	MILLHLDFLS	ALLYAAVFLF	LIFRAGMLQW	FWASIMLWLG	ISVLGAKLMP
30	51	GIWGMTRAAP	LEIPHFYLT	GSIFFFIGHW	NRKTDGNGWQ	ADPEHPLLGL
	101	FAVSNVSMTL	AFVGICALVH	YCFSTVQVVF	VFAALLKLYA	LKPVYWFVLQ
	151	FVLMAYVYVH	RCGIDRQPPS	TFGGSQRLRLG	GLTAALMQXS	VLVLLVLLSEI
	201	R*				

ORF18a and ORF18-1 show 99.0% identity in 201 aa overlap:

		10	20	30	40	50	60
35	orf18a.pep	MILLHLDFLSALLYAAVFLF	LIFRAGMLQWFWASIMLWLG	ISVLGAKLMPGIWGMTRAAP			
	orf18-1	MILLHLDFLSALLYAAVFLF	LIFRAGMLQWFWASIMLWLG	ISVLGAKLMPGIWGMTRAAP			
		10	20	30	40	50	60
40	orf18a.pep	LFIPHFYLTGSIFFIGHWNRKTDGNGWQADPEHPLLGLFAVSNVSMTLAFVGICALVH					
	orf18-1	LFIPHFYLTGSIFFIGHWNRKTDGNGWQADPEHPLLGLFAVSNVSMTLAFVGICALVH					
		70	80	90	100	110	120
45	orf18a.pep	YCFSTVQVVFVAALLKLYALKPVYWFVLQFVLMAYVYVHRCGIDRQPPSTFGGSQRLRLG					
	orf18-1	YCFSTVQVVFVAALLKLYALKPVYWFVLQFVLMAYVYVHRCGIDRQPPSTFGGSQRLRLG					
		130	140	150	160	170	180
50	orf18a.pep	GLTAALMQXS	VLVLLVLLSEIGRX				
	orf18-1	GLTAALMQXS	VLVLLVLLSEIGRX				
		190	200				
55	orf18a.pep	GLTAALMQXS	VLVLLVLLSEIGRX				
	orf18-1	GLTAALMQXS	VLVLLVLLSEIGRX				
		190	200				

Homology with a predicted ORF from *N.gonorrhoeae*

ORF18 shows 93.1% identity over a 116aa overlap with a predicted ORF (ORF18.ng) from *N.*

60 *gonorrhoeae*:

	orf18.pep	GNGWQADPEHPLLGLFAVSNVSMTLAFVGI	30
	orf18ng	TRAAPLFIPHFYLTLSGIFFFIYWNKRKTGNGWQADPEHPLLGLFAVSNVSMTLAFVGI	115
5	orf18.pep	CALVHYCFSGTVQVVFVFAALLKLYALKPVYWFVLQFVLMAYAVVHRCGIDRQPPSTFGGS	90
	orf18ng	CALVHYCFSGTVQVVFVFAALLKLYALKPVYWFVLQFVLMAYAVVHRCGIDRQPPSTFGGS	175
10	orf18.pep	QLRLGGLTAALMQVSVLVLLSEIGR	116
	orf18ng	QLRLGVLAAMLQVAVTAMLLAEIGR	201

The complete length ORF18ng nucleotide sequence is <SEQ ID 101>:

1	ATGATTTCG	TGCATTTCGA	TTTTTGTCT	GCCTTACTGt	aTGCgGcggt
15	51	ttttctctgTTT	CTGATATTC	CGCAGGAAT	GTTCGAATGG
	101	GTATTTCGCTT	GTGCTCTGCG	ATCTCGGTT	TAGGGGTAAA
	151	GGGATGTGGG	GATGACCCG	CGCCGCGCT	TTCTTCATCC
	201	CTGACCTTG	GGCAGCATAT	TTTTTTTCAT	CGGGTATTGG
	251	CAGATGGAAA	CGGATGCGCAG	CGAGACCCG	AACATCCGCT
20	301	TTTCCCGTCA	GTAATGTATC	GATGACGCTT	GCTTTGTGCG
	351	GTTGGTGCAT	TATTGCTTTT	CGGGAACGGT	TCAAGTGTTT
	401	CATTGCTCAA	ACTTTATGCG	CTGAAGCCGG	TTTATTGGTT
	451	TTTGTATTGA	TGGCGGttgC	CTATGTCCAC	CGCTCGCGTA
	501	CGCCCGGTCA	ACGTTGCGCG	GTTGCGAGCT	CGCACTCGGC
25	551	CGATGTTGAT	GCAGGTTGCG	GTAACGCGGA	TGCTGCTTGC
	601	AGATGA			

This encodes a protein having amino acid sequence <SEQ ID 102>:

1	MILLHLDFLS	ALLYAAVFLF	LIFRAGMLQW	FWASIALWLG	ISVLGVKLMP
30	51	GMWGMTRAAP	LFIPHFYLT	LSIFFFIYWN	NKRKTGNGWQ
	101	FAVSNVSMTL	AFVGICALVH	YCFSGTVQVF	VFAALLKLYA
	151	FVLMAYAVVH	RCGIDRQPPS	TFGGSQRLG	VLAAMLQVAV
	201	R*			

This ORF18ng protein sequence shows 94.0% identity in 201 aa overlap with ORF18-1:

35	orf18-1.pep	10	20	30	40	50	60
	orf18ng	10	20	30	40	50	60
40	orf18-1.pep	70	80	90	100	110	120
	orf18ng	70	80	90	100	110	120
45	orf18-1.pep	130	140	150	160	170	180
	orf18ng	130	140	150	160	170	180
50	orf18-1.pep	190	200				
	orf18ng	190	200				

Based on this analysis, including the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 13

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 103>:

```

      1  ATGAAAACCC  CACTCCTCAA  GCCTCTGCTN  ATTACTCTGC  TTCCCGTTTT
5    51  CGCCAGTGTT  TTTACCGCGC  CCTCCATCGT  CTGGCAGCTA  GGCGAACCCTA
      101  AGCTCGGCAT  GCCTTCGTGA  CTGCGCATCA  TCGGCGGCGG  CCTTGTGCGAT
      151  TTGGACAACC  NCGTACCCTG  ACGGCTNAAA  AACATCATCA  CCACGCTCGC
      201  CCGTGTTCACC  CTCTCCTCGC  TCACGGCACA  AAGCACCCCT  GGCACAGGGC
      251  TGCCCTTCAT  CCTGCCATG  ACCCTGATGA  CTT.CG.CTT  CACCATTTTA
      301  GGCGCGNGC  ...

```

10 This corresponds to the amino acid sequence <SEQ ID 104; ORF19>:

```

      1  MKTPLLKPLL  ITSLEVFASV  FTAASIVQWL  GEPKLAMPFV  LGIIAGGLVD
      51  LDNXXTGRKL  NIITVALFT  LSSLTAQSTL  GTGLPFILAM  TLMTXXFIL
     101  GAX...

```

Further work revealed the complete nucleotide sequence <SEQ ID 105>:

```

15    1  ATGAAAACCC  CACTCCTCAA  GCCTCTGCTC  ATTACTCTGC  TTCCCGTTTT
      51  CGCCAGTGTT  TTTACCGCGC  CCTCCATCGT  CTGGCAGCTA  GGCGAACCCTA
     101  ASCTCGGCAT  GCCTTCGTGA  CTGCGCATCA  TCGGCGGCGG  CCTTGTGCGAT
     151  TTGGACAACC  GCCTGACCGG  ACGGCTGAAA  AACATCATCA  CCACGCTCGC
     201  CCGTGTTCACC  CTCTCCTCGC  TCACGGCACA  AAGCACCCCT  GGCACAGGGC
     251  TGCCCTTCAT  CCTGCCATG  ACCCTGATGA  CTT.CG.CTT  CACCATTTTA
     301  GGCGCGNGC  ...
35    351  CGCCAGTGTT  TTTACCGCGC  CCTCCATCGT  CTGGCAGCTA  GGCGAACCCTA
     401  AGCTCGGCAT  GCCTTCGTGA  CTGCGCATCA  TCGGCGGCGG  CCTTGTGCGAT
     451  TTGGACAACC  GCCTGACCGG  ACGGCTGAAA  AACATCATCA  CCACGCTCGC
     501  CCGTGTTCACC  CTCTCCTCGC  TCACGGCACA  AAGCACCCCT  GGCACAGGGC
     551  TGCCCTTCAT  CCTGCCATG  ACCCTGATGA  CTT.CG.CTT  CACCATTTTA
     601  GGCGCGNGC  ...
40    651  CGCCAGTGTT  TTTACCGCGC  CCTCCATCGT  CTGGCAGCTA  GGCGAACCCTA
     701  AGCTCGGCAT  GCCTTCGTGA  CTGCGCATCA  TCGGCGGCGG  CCTTGTGCGAT
     751  TTGGACAACC  GCCTGACCGG  ACGGCTGAAA  AACATCATCA  CCACGCTCGC
     801  CCGTGTTCACC  CTCTCCTCGC  TCACGGCACA  AAGCACCCCT  GGCACAGGGC
     851  TGCCCTTCAT  CCTGCCATG  ACCCTGATGA  CTT.CG.CTT  CACCATTTTA
     901  GGCGCGNGC  ...
45    951  CGCCAGTGTT  TTTACCGCGC  CCTCCATCGT  CTGGCAGCTA  GGCGAACCCTA
     1001  AGCTCGGCAT  GCCTTCGTGA  CTGCGCATCA  TCGGCGGCGG  CCTTGTGCGAT
     1051  TTGGACAACC  GCCTGACCGG  ACGGCTGAAA  AACATCATCA  CCACGCTCGC
     1101  CCGTGTTCACC  CTCTCCTCGC  TCACGGCACA  AAGCACCCCT  GGCACAGGGC
     1151  TGCCCTTCAT  CCTGCCATG  ACCCTGATGA  CTT.CG.CTT  CACCATTTTA
     1201  GGCGCGNGC  ...
50    1251  CGCCAGTGTT  TTTACCGCGC  CCTCCATCGT  CTGGCAGCTA  GGCGAACCCTA
     1301  AGCTCGGCAT  GCCTTCGTGA  CTGCGCATCA  TCGGCGGCGG  CCTTGTGCGAT
     1351  TTGGACAACC  GCCTGACCGG  ACGGCTGAAA  AACATCATCA  CCACGCTCGC
     1401  CCGTGTTCACC  CTCTCCTCGC  TCACGGCACA  AAGCACCCCT  GGCACAGGGC
     1451  TGCCCTTCAT  CCTGCCATG  ACCCTGATGA  CTT.CG.CTT  CACCATTTTA
     1501  GGCGCGNGC  ...
55    1551  CGCCAGTGTT  TTTACCGCGC  CCTCCATCGT  CTGGCAGCTA  GGCGAACCCTA
     1601  AGCTCGGCAT  GCCTTCGTGA  CTGCGCATCA  TCGGCGGCGG  CCTTGTGCGAT
     1651  TTGGACAACC  GCCTGACCGG  ACGGCTGAAA  AACATCATCA  CCACGCTCGC
     1701  CCGTGTTCACC  CTCTCCTCGC  TCACGGCACA  AAGCACCCCT  GGCACAGGGC
     1751  TGCCCTTCAT  CCTGCCATG  ACCCTGATGA  CTT.CG.CTT  CACCATTTTA
     1801  GGCGCGNGC  ...
     1851  CGCCAGTGTT  TTTACCGCGC  CCTCCATCGT  CTGGCAGCTA  GGCGAACCCTA
     1901  AGCTCGGCAT  GCCTTCGTGA  CTGCGCATCA  TCGGCGGCGG  CCTTGTGCGAT
     1951  TTGGACAACC  GCCTGACCGG  ACGGCTGAAA  AACATCATCA  CCACGCTCGC
     2001  CCGTGTTCACC  CTCTCCTCGC  TCACGGCACA  AAGCACCCCT  GGCACAGGGC
     2051  TGCCCTTCAT  CCTGCCATG  ACCCTGATGA  CTT.CG.CTT  CACCATTTTA
     2101  GGCGCGNGC  ...
     2151  A

```

This corresponds to the amino acid sequence <SEQ ID 106; ORF19-1>:

```

60    1  MKTPLLKPLL  ITSLEVFASV  FTAASIVQWL  GEPKLAMPFV  LGIIAGGLVD

```

5

10

15 Computer analysis of this amino acid sequence gave the following results:

Homology with predicted transmembrane protein YHFK of *H. influenzae* (accession number P44289)

ORF19 and YHFK proteins show 45% aa identity in 97 aa overlap:

20

25

```

orf19 6  LKPLLITSLPVFASVFTAASIVWQLGEPKLPAMPFVLGIAGGLVDLDNXXTGRLKNIIIT 65
      L +I++++PVF +V AA +W +MP +LGIAGGLVDLDN TGRLEK+ T
YHFK 5  LNAKVISTIPVFIAVNIAAVGIWFFDISQSMPLLGIAGGLVDLDNRLTGRLEK 64

orf19 66 VALFTLSSLTAAQSTLGTGLPFLIAMTLMXXFTILGA 102
      + F++SS Q +G + +I+ MT++T FT++GA
YHFK 65 LIAFSISSPFVQLHIGKEFIQYIVLMTVLFFIFMIGA 101

```

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF19 shows 92.2% identity over a 102aa overlap with an ORF (ORF19a) from strain A of *N.*

meningitidis:

30

35

40

```

orf19.pep 10 20 30 40 50 60
          MKTPLLKPLLITSLPVFASVFTAASIVWQLGEPKLPAMPFVLGIAGGLVDLDNXXTGRLEK
          |||||
orf19a     10 20 30 40 50 60
          MKTPEPLKPLLITSLPVFASVFTAASIVWQLGEPKLPAMPFVLGIAGGLVDLDNRLTGRLEK

orf19.pep 70 80 90 100
          NIIITVALFTLSSLTAAQSTLGTGLPFLIAMTLMXXFTILGAX
          |||||
orf19a     70 80 90 100 110 120
          NIIATVALFTLSSLTAAQSTLGTGLPFLIAMTLMFGFTIMGAVGLKYRTFAFGALAVATY

orf19a     130 140 150 160 170 180
          TTIITYTEPTYWLNTNPFMILCGTVLYSTAILFQIILPHRPVQENVANAYEALGSYLEAKA

```

The complete length ORF19a nucleotide sequence <SEQ ID 107> is:

45

50

55

```

1  ATGAAACCC CACCCCTCAA GCCTCTGCTC ATTACCTGCG TTCCCGTTTT
5  CGCCAGTGTC TTTACGCGCG CCTCATCGCT CTGGCAGCTG GCGCAACCCA
10  AGCTCGCCAT GCCTCTTGCTA CTGCGCATCA TCGCTGGCGG CCTGGTGCAT
15  TTGGACAACC GCCTGACCGG ACGGCTGAAG AACATCATCG CCACGCTGCG
20  CTGTTCACCC CTCTCCTCAT TTGTGCGCGA AAGCACCGCT GGCACAGGTT
25  TGCATCTATT CCTGCGCATG ACCGTATGTA CTTCGCGGCT TACCATCATG
30  GCGCGGCTCG GCGTGAANTA CCGCACCTTC GCCTCTGCGG CACTCGCGCT
35  CGCCACCTAC ACCACACTTA CCTACACCCC CGAAACCTAC TGGCTGACCA
40  ACCCTTTTAT GATTCTGTGC GGAACCGTAC TGTACAGCAC CGCCATCATC
45  CTGTTCCAAA TCATCCTGCC CCACGCGCCC GTTCAAGAAA ACGTCGCCAA
50  CGCCTACGAA GCACTCGGCA GCTACCTCGA AGCCAAGGCC GACTTTTTCG
55  ATCCCGACGA AGCCGAATGG ATAGGCAACC GCCACATCGA CCTCGCCATG
60  AGCAACACCG CGCTCATCAC CGCCTTCAAC CAATGCGGTT CGCGCCTGTT
65  TTACCGCCTT CGCGGCAACC ACCGCCACCC CGCAGCCGCG AAAATCGTGC
70  GCTACTACTT CGCCGCCCAA GACATACAGC AAGCATCAG CTCCGCCAC
75  GTCGACTACC AAGAGATGTC CGAAAAATTC AAAAACACCG ACATCATCTT

```

801 COGCATCCAC GCCTGCTCG AATGCAGGG ACAAGCCTGC CGCAACACGG
 851 CCCAGAGCTT GCGCGCAAGC AAGACTACG TTTACAGCAA ACGCCCTCGG
 901 CGCGCCATCG AAGGCTGCCG CAAATCGCTG GCCTCTCTTT CACAGACGAA
 951 CGCAGATCCC GACATCCSCG ACCTCGCGCG CTTTCTCGAC AACCTCGGCA
 1001 GCGTCGACCA CAGATTCCGC CAACTCCAGC ACACGCGCCT CGAGGCGAAA
 1051 AAGGACCGCA TGGGCGACAC CCGCATCGCC GCCTCGAAA CCGCGAGCCT
 1101 CAAAACACCC TGGCAGGCAA TCGCTCCGCA GCTAAACCTC GAATCAGGCG
 1151 TATTTCCGCC TGCCTGCTCG CTGTCCCTTG TCGTTGCCCG CGCTGCAACC
 1201 ATGCTCGAAG CCTCAACCT CAACTCGGCG TACTGGATAC TACTGACCGC
 1251 CCTTTTCGTG TGCCAAACCA ACTACACCGC CACCAAAAGC CGCGTCCGCC
 1301 AGCGCATCGC CGGCAACGTA CTCGGCGTAA TCGTGGCTCC GCTGCTCCCC
 1351 TACTTTACCC CCTCGCTCGA AACCAACTC TGGATCGTCA TCGCCAGTAC
 1401 CACCCTCTTT TTATGACCC GCACCTACAA ATACAGCTTC TCGACATTTT
 1451 TCATCAACAT TCAAGCCCTG ACCAGCCTCT CCCTCGCAGG GTTSGACGTA
 1501 TAGCGCGCCA TGCCCGTAGC CATCATCGAC ACCATTATCG CGGCATCCCT
 1551 TGCTTGGGCG CGAGTCAGCT ACCTGTGGCG AGACTGGAAA TACTCACGCG
 1601 TCGAAGCGAC GAGCGCCCTT GCGCTATGCA GCAACGCGCG CTATCTCGAA
 1651 AAAATCAGCC AAGCGCTCAA AAGCGCGCAA ACCGCGACG ACTGCGATA
 1701 CCOCGCCACG CGCCGCCGCG CCACGAAACA CACCGCGCCG CTCAGCAGCA
 1751 CCCTTTCCGA CATGAGCAGC GAACCCGCAA AATTCCGCGA CAGCCTGCAA
 1801 CCGCGCTTTA CCTGCTCAA AACCGGCTAC GCCTGACCG GCTACATCTC
 1851 CGCCCTCGCG CATAACGCA GCGAATATGA CGAAGAATCG AGCCCCAGCT
 1901 TTACCGCACA GTTCCACCTC GCGCCGGAAC ACACGCGCCA CATCTTCAA
 1951 CACCTGCCCG AAACCGAACC CGACGACTTT GACACGACAT TGGATACACT
 2001 GCGCGGCGAA CTGCAACCCC TCCGACCCCA CAGCAGCGGA ACACAAAGCC
 2051 ACATCTCTCT CCAACAGCTC CAACTCATCG CCAGCGCAGT CGAACCTCTAC
 2101 TACCGCGCCT ACCGACAAAT TCCGCACAGG CAGCCCCAAA ACGCAGCCTG
 2151 A

This encodes a protein having amino acid sequence <SEQ ID 108>:

1 MKTPPLKPLD ITSLPVFASV FTAASIVWQL GEPKLAMPFV LGIIAGGLVD
 51 LDNRLTGRLL NIATVALFT LSSLVAQSTL GTGLPFILAM LTIITFGFTIM
 101 GAVGLKYRTP AFGALAVATY TLTLYTPETY WLTNPFMILC GIVLYSTAIL
 151 LFQIILPHRP VQENVANAYE ALGSGYLEAKA DFDPDEAEW IGKRRHDLAM
 201 SNTGVITAFN QGRSLAEVRL RKGKRRHPTA DIERHTSSAH
 251 VDYQEMSEKF KNTDLIERIH RLEMQGQAC RNTAQLARAS KDYYYSKRLG
 301 RATEGCROSL RLLSDSNNDP DIRHLRLLLD NLGSVDQQFR QIQHNGLOAE
 351 NDRMGDTRIA ALETGSLKNT WQAIRFQNLN ESGVFRHAVR LSLVVAACAT
 401 IVEAINLNLG YWILLTALFV QPNYATKSS RVRQRIAGTV LGVIVGSLVP
 451 YTFPSVETKL WIVIASTTLF FMRTYKYSF TFFITIOAL TSLSLAGLDV
 501 YAAMPVRIID TIIGASLAWA AVSYLWPDWK YLTLEITAAI AVCSNGAYLE
 551 KITERLKSGE TGDDVEYRAT RRRRAHTAAS LSTSLDMSS EPAKFADSLQ
 601 PGFTLLKTGY ALTGYISALG AYRSEMHEEC SPDFTAQFHL AAEHTAHIFQY
 651 HLPETEPDDF QTALDTRIGE LDTLRTHSSG TQSHILLQOL QLIARQLEPY
 701 YRAYRQIPHR QPQNA*

ORF19a and ORF19-1 show 98.3% identity in 716 aa overlap:

		10	20	30	40	50	60
orf19a.pep		MKTPPLKPLDITSLPVFASVFTAASIVWQLGEPKLAMPFV	LGIIAGGLVD	LDNRLTGRLL			
orf19-1		MKTPPLKPLDITSLPVFASVFTAASIVWQLGEPKLAMPFV	LGIIAGGLVD	LDNRLTGRLL			
		10	20	30	40	50	60
		70	80	90	100	110	120
orf19a.pep		NIATVALFTLSSLVAQSTLTGTGLPFILAMLTMTFGFTIM	GAVGLKYRTPAFGALAVATY				
orf19-1		NIITVALFTLSSLVAQSTLTGTGLPFILAMLTMTFGFTIL	GAVGLKYRTPAFGALAVATY				
		70	80	90	100	110	120
		130	140	150	160	170	180
orf19a.pep		TTLTYPETYWLTNPFMILCGTVLYSTAILLFQIILPHRP	VQENVANAYEALGSGYLEAKA				
orf19-1		TTLTYPETYWLTNPFMILCGTVLYSTAILLFQIILPHRP	VQENVANAYDALGGYLEAKA				
		130	140	150	160	170	180
		190	200	210	220	230	240
orf19a.pep		DFDPDEAEWIGNRHLDMASNTGVITAFNQCRSALFYRL	RKGKRRHPTAKMLRYFAAQ				

	orf19-1	DDFFDPDEAAWIGNRHIDLAMSNTGVITAFNQCRSALFYRLRGKRRHPTAKMLRYFYFAAQ	190	200	210	220	230	240
			250	260	270	280	290	300
5	orf19a.pep	DIHERISSAHVDYQEMSEKFKNTDIIIFRIHRLLEMQGQACRNTAQALRASKDYYVSKRLG						
	orf19-1	DIHERISSAHVDYQEMSEKFKNTDIIIFRIHRLLEMQGQACRNTAQALRASKDYYVSKRLG	250	260	270	280	290	300
10	orf19a.pep	RAIEGCRQSLRLLSDSNNDPDIRHLRRLLDNLGSDVDQFRLQHNGLQAENDRMGDTRIA	310	320	330	340	350	360
	orf19-1	RAIEGCRQSLRLLSDSNDSFDIRHLRRLLDNLGSDVDQFRLQHNGLQAENDRMGDTRIA	310	320	330	340	350	360
15	orf19a.pep	ALETGSLKNTWQAIRPQNLLESGVFRHAVRLSLVVAACACTIVEALNINLGYWILLTALFV	370	380	390	400	410	420
	orf19-1	ALETGSLKNTWQAIRPQNLLESGVFRHAVRLSLVVAACACTIVEALNINLGYWILLTALFV	370	380	390	400	410	420
20	orf19a.pep	CQFNATATKSRVRQRIAGTVLGVIVGSLVPYFTPSVETKLWIVIASTTLFFMTRTYKYSF	430	440	450	460	470	480
	orf19-1	CQFNATATKSRVRQRIAGTVLGVIVGSLVPYFTPSVETKLWIVIASTTLFFMTRTYKYSF	430	440	450	460	470	480
25	orf19a.pep	STFFITIQALTSLSLAGLDVYAAMPVRIIDTIIIGASLAWAAVSYLWPDWKYLTLERTAAL	490	500	510	520	530	540
	orf19-1	STFFITIQALTSLSLAGLDVYAAMPVRIIDTIIIGASLAWAAVSYLWPDWKYLTLERTAAL	490	500	510	520	530	540
30	orf19a.pep	AVCSNGAYLEKITERLKSGETGDDVEYRATRRRAHEHTAALSSTLSDMSSPAKFADSLQ	550	560	570	580	590	600
	orf19-1	AVCSNGAYLEKITERLKSGETGDDVEYRATRRRAHEHTAALSSTLSDMSSPAKFADSLQ	550	560	570	580	590	600
35	orf19a.pep	PGFTLLKTYGALTGYISALGAYRSEMHEECSPDFTAQFHLLAAEHTAHIFQHLPETEPDDF	610	620	630	640	650	660
	orf19-1	PGFTLLKTYGALTGYISALGAYRSEMHEECSPDFTAQFHLLAAEHTAHIFQHLPETEPDDF	610	620	630	640	650	660
40	orf19a.pep	QTALDTLRGELDTLRTHSSQTSQHILLQQLIARQLEPYRAYRQIPHRCQNAAX	670	680	690	700	710	
	orf19-1	QTALDTLRGELDTLRTHSSQTSQHILLQQLIARQLEPYRAYRQIPHRCQNAAX	670	680	690	700	710	

Homology with a predicted ORF from *N.gonorrhoeae*

ORF19 shows 95.1% identity over a 102aa overlap with a predicted ORF (ORF19.ng) from *N. gonorrhoeae*:

55	orf19.pep	MKTPLLKPLLITSLPVFASVFTAASIVWQLGEPKLMFPVLGIIAGGLVDLNDXXTGRLK	60
	orf19ng	MKTPLLKPLLITSLPVFASVFTAASIVWQLGEPKLMFPVLGIIAGGLVDLNDLRTGRLK	60
60	orf19.pep	NIITVALFTLSLTAQSTLTGTLPFILAMTMTXXFTILGAX	103
	orf19ng	NIITVALFTLSLTAQSTLTGTLPFILAMTMTFGFTILGAVGLKYRTFAFGALAVATY	120

An ORF19ng nucleotide sequence <SEQ ID 109> is predicted to encode a protein having amino acid sequence <SEQ ID 110>:

1 MKTPLLKPLL ITSLLPVFASV FTAASIVWQL GEPLAMPFV LGIIAGGLVD
 51 LDNRLTGRLK NIIATVALFT LSSLTAQSTL GTGLPFILAM TLMTFGFITL
 101 GAVGLKYRTF AFGALAVATY TLTLYTPETY WLTNPFMILC GTVLYSTAIL
 151 LFOIILPHRE VQESVANAYE ALGGYLEAKA DFFDPDEAAW IGRNRHIDLAM
 5 SNTGVITAFN QCRSALFYRL RGKRRHPRTA KMLRYFFAAQ DIHERISSAH
 251 VDYQEMSEKF KNTDIIIFRIR RLLEMQGQAC RNTAQAIRSG KDYYYSKRLG
 301 RATEGCRQSL RLLSDGNDSP DIRHLSRLLD NLGSVDQQFR QLRHSDSPAE
 351 NDRMGDTRIA ALETGSFKNT *

Further work revealed the complete nucleotide sequence <SEQ ID 111>:

1 ATGAAACCC CACTCCTCAA GCCTCTGCTC ATTACCTCGC TTCCCGTTTT
 51 CGCCAGTGTC TTTACGCCGG CCTCATCGT CTGGCAGCTA GCGCAACCCA
 101 AGCTCGCCAT GCCCTTCGTA CTGGCAGTCA TCGCCGGCGG CCTGTCGAT
 151 TTGGACAACC GCCTGACCGG ACGGCTGAAA AACATCATCG CCACCGTCGC
 201 CCTGTTTACC CTCTCCTCGC TCACGGCGCA AAGCACCCCT GGCACAGGCG
 15 TGCCCTTCAT CCTCGCATG ACCCTGATGA CCTTCGGCTT TACCATTTTA
 301 GGGCGGGTGC GGCTGAAATA CCGCACCTTC GCCTTCGGCG CACTCGCGGT
 351 CGCCAGCTAC ACCAGCCTTA CCTACACCCC GAAACCTTAC TGGCTGACCA
 401 ACCCCTTCAT GATTTAATGC GGACCGCTAC TGTACAGCAC GGCATCATCA
 451 CTGTTCCAAA TCAATCCTCGC CCACCGCCCG CTCCAGAAAG CGCTGCCCAA
 501 TGCCCTACGAA GCACTCGGGC GCTACCTCGA AGCCCAAAGC GCTCTTCTCG
 551 ACCCGCATGA GGCAGCTGG ATAGCGCAAC GCCACATCGA CCTCGCCATG
 601 AGCAACACCG GCGTCATCAC GCCTTCAAC CAATGCCGTT CGCGCCTGTT
 651 TTACCGTTTG CGCGCAACAC ACCGCCACCC GGGCAGCCGC AAAATGCTGC
 701 GCTACTACTT CGCCGCCCAA GACATCCAG AGCAGCATCAG CTGCGCCAC
 25 751 GTGCACTACC AAGAGATGTC CGAAAAATTC AAAACACCGC ACATCATCTT
 801 CGCGATCCGC GCCTGCTCGC AAATGCAAGG CAGCGGTGCG CGCAACACCG
 851 CCCAAGCCAT CGGTCGGCG AAAGACTAG tTTACGCAA ACGCCCTCGA
 901 CGCGCCATCG aaggtctCGC CAGTCTGCTg cgctcCTTT cagacggcaa
 951 CGACAGTCCC GACATCCGCC ACCTGAGCGg CTTCTTCGAC AACCTCGcaa
 30 1001 GCGTCgacaa cgagtTCcgc caactCCGAC ACAGcgactc CCGCGcgaa
 1051 Aacgacgcga tggggcgaca CCGCATCGCC GCCCTcgaaa cggcgagctt
 1101 caaaaaCacc tggcagggCAA TCGCTGCTCA gctgaacCTC GAATgactCG
 1151 TATTTCCTCA CTGTCCTCGC CTGTCCTCGC TCGTTCGCGC CGCGCTCGCA
 1201 ATCGTCgaag cCTCAACCT CAACCTCGGC TACTGTGATC TCGTGACCGC
 35 1251 CTTTTCCTC TGCCAAACCA ACTACACCGC CACCAAAGC CGGCTGTACC
 1301 AACGCATCGC CGGCACCGTA CTCGCGGTAA TCGTGGCTCG GCTGTCGCC
 1351 TACTTCAACC CCTCCGTGCA AACCAAACTC TGGATTGCA TCGCGCGTAC
 1401 CACCTGTGTC TTCATGACCC GCACCTACAA ATACAGTTTC TCCACCTTCT
 1451 TCATCACCAT TCAGGCACGT ACCAGCCTCT CCTCGCAGG TTGGACGTA
 40 1501 TACGCGCGCA TGCCGCTGCG CATCATgcaC ACCATTATCG GCGCATCCTT
 1551 TGCCCTGGCG GCGGTCACTT ACCTGTGGCC AGACTGGAAA TACCTCAGCG
 1601 TCGAACGCAC CGCGCCCTT GCCGTATGCA CGACGGGCAC ATACCTCCAA
 1651 AAAATTGCGC AAGCCTCAA AACCGCGGAA ACCGGCGGAC ACATAGATA
 45 1701 CGCATCACCC CGCGCGCGG CCCACGAACA CACCGCGGCC CTCAGCAGA
 1751 CCGTTTCCGA CATGAGCAGC GAACCGGCAA AATTCCGCGA CAGCGTGCAA
 1801 CCGCGCTTTA CCTGTCTCAA AACCGGCTAC CGCCTGACCG GCTACATCTC
 1851 CGCCCTCGGC GCATACCGCA CGGAATGCA CGAGAATGC AGCCCGCAT
 1901 TTACCGCACR GTTCACCTT GCAGCGGAC ACACCGGCGA CATCTTCAA
 1951 CACTCGCGC ACATGGAGAC CAGCAGCTT CAGACGCTAT TGGATACCT
 50 2001 CGCGCGCGAA CTCGCGACCC TCCGACCCCG CAGCAGCGCA ACACAGCC
 2051 ACATCCTCCT CCAACAGCTC CAATCATCGC CCGGCGMACT CGAACCTTAC
 2101 TACCGCGCCT ACCGACAAAT TCCGACACAG CGGACCCAAA ACGCAGCCTG
 2151 A

This corresponds to the amino acid sequence <SEQ ID 112; ORF19ng-1>:

1 MKTPLLKPLL ITSLLPVFASV FTAASIVWQL GEPLAMPFV LGIIAGGLVD
 51 LDNRLTGRLK NIIATVALFT LSSLTAQSTL GTGLPFILAM TLMTFGFITL
 101 GAVGLKYRTF AFGALAVATY TLTLYTPETY WLTNPFMILC GTVLYSTAIL
 151 LFOIILPHRE VQESVANAYE ALGGYLEAKA DFFDPDEAAW IGRNRHIDLAM
 60 SNTGVITAFN QCRSALFYRL RGKRRHPRTA KMLRYFFAAQ DIHERISSAH
 251 VDYQEMSEKF KNTDIIIFRIR RLLEMQGQAC RNTAQAIRSG KDYYYSKRLG
 301 RATEGCRQSL RLLSDGNDSP DIRHLSRLLD NLGSVDQQFR QLRHSDSPAE
 351 NDRMGDTRIA ALETGSFKNT WQAIRPQHV RSCVFHRAV LSLVVAACAT
 401 IVEALNLLG WYLLTALFY CQRYATKPS RYVQIRAGTY LGVTVGSIVE
 451 YTFPSVETKL WIVIAGTTFL VFSRYTKYIS STFTITIOAL TSLSLAGL
 65 501 YAMAPVRID TIIGASLAWA RVSYLAPDQWK YLTERTAAAL AVCSGTYLQ
 551 KIAERLKTGE TGDDBIYRIT RRAHEHTAA LSLTSLDMSM EFAKFADSLQ

		610	620	630	640	650	660
		670	680	690	700	710	
5	orf19-1.pep	QTALDTRLRGELDTLRTHSSGTQSHILLQQLQLIARQLEPPYRAYRQIPHRQPQNAAX					
	orf19ng-1	QTALDTRLRGELGTLRTRSSGTQSHILLQQLQLIARQLEPPYRAYRQIPHRQPQNAAX					
		670	680	690	700	710	

In addition, ORF19ng-1 shows significant homology to a hypothetical gonococcal protein previously entered in the databases:

10	sp O33369 YOR2_NEIGO_HYPOTHETICAL_45.5_KD_PROTEIN_(ORF2)_gnl PID e1154438 (AJ002423) hypothetical protein [Neisseria gonorrh] Length = 417 Score = 1512 (705.6 bits), Expect = 5.3e-203, P = 5.3e-203 Identities = 301/326 (92%), Positives = 306/326 (93%)
15	Query: 307 RQSLRLSDGNDSPDIRHLSRLLDNLGSDVQDQFRQLRHSDSPAENDRMGDTRIAALETGS 366 RQSLRLSDGNDSDIRHLSRLLDNLGSDVQDQFRQLRHSDSPAENDRMGDTRIAALETGS Sbjct: 1 RQSLRLSDGNDSDXDIRHLSRLLDNLGSDVQDQFRQLRHSDSPAENDRMGDTRIAALETGS 60
20	Query: 367 FKNTWQAIKQNLNLESCVFRHVAVRLSLVVAACCTIVEALNINLGYWILLTALFVQCPNYT 426 FKNTWQAIKQNLNLESCVFRHVAVRLSLVVAACCTIVEALNINLGYWILLTALFVQCPNYT Sbjct: 61 FKNTWQAIKQNLNLESCVFRHVAVRLSLVVAACCTIVEALNINLGYWILLTALFVQCPNYT 120
25	Query: 427 ATKSRVYQRIAGTVLGVIVGSLVPYFTPSVETKLWIVAGTTLFFMRTYKYFSSTFFIT 486 ATKSRVYQRIAGTVLGVIVGSLVPYFTPSVETKLWIVAGTTLFFMRTYKYFSSTFFIT Sbjct: 121 ATKSRVYQRIAGTVLGVIVGSLVPYFTPSVETKLWIVAGTTLFFMRTYKYFSSTFFIT 180
30	Query: 487 IQALTSLSLAGLDVYAAMPVRIIDTIIGASLAWAAVSYLWPDWKYLTERTAAALAVCSSG 546 IQALTSLSLAGLDVYAAMPVRIIDTIIGASLAWAAVSYLWPDWKYLTERTAAALAVCSSG Sbjct: 181 IQALTSLSLAGLDVYAAMPVRIIDTIIGASLAWAAVSYLWPDWKYLTERTAAALAVCSSG 240
35	Query: 547 TYLQKIAERLKTGETGDDIEYRITRRRAHEHTAALSTSLSDMSSEPAKFADSLQPGFTLL 606 TYLQKIAERLKTGETGDDIEYRITRRRAHEHTAALSTSLSDMSSEPAKFADSLQPGFTLL Sbjct: 241 TYLQKIAERLKTGETGDDIEYRITRRRAHEHTAALSTSLSDMSSEPAKFADSLQPGFTLL 300
	Query: 607 KTGYSALGYSALGAYRSMHEECSP 632 K ALTGYSALG ++ + +P Sbjct: 301 KPATALGYSALGHTAAKCTKNAAP 326

Based on this analysis, including the presence of several putative transmembrane domains in the gonococcal protein (the first of which is also seen in the meningococcal protein), and on homology with the YHFK protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 14

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 113>:

45	1	ATGAATAATGC	TGGGAGCTTT	GGCAAAAGTC	GGCAGCCTGA	CGATGGTGTC
	51	GCGCGTTTTTG	GGATTITGTGC	CGGATACGGT	CATTGCGCGG	GCATTCCGGC
	101	CGGGTATGCG	GACGGATGCG	TTTTTTGTGC	CGTTCAAACCT	GCCCAACCTG
	151	CTTCGCGCGG	TGTTTTCGGGA	GGGGGCGGTT	GCCCAAGCGT	TTGTGCGCAT
	201	TTTGGCGGAA	TACAAGGAAA	CGCGTTCAAA	AGAGCGCGG	CAAGACCTTTA
50	251	TCCGCCATGT	GCGCGGGATG	CTGTGCTTTG	TACTGGTTAT	CGTTACCGCG
	301	CTGGGCGATAC	TTGCGCGGCC	TTGGGTGATT	TATGTTTCCG	CACCGCGATT
	351	TTGCCCAAGA	TGCGGACAAA	TTTCAGCTCT	CATCGATTCT	GCTTCGAGATT
	401	ACGTTTCCTT	ATATATTTAT	GATTTCCTGT	TCTTCATTGT	TCGCGTCGGT
	451	ACTCAATTCT	TATCATATGT	TGCGCATATG	CGAGGTTC	GTTC
55	501	TGAACGTGTC	GTTTATCGTA	TTCCGCTGCT	TTTTCGTGCC	GTATTTCGAT

551 CCGCCCGTTA CCGGCyGGC GTGGGCGGTC TTTGTCGGCG GCATTTTGA
 601 ACTCGrmtTC CAACTGCCCT GGCTGGCGAA ACTGGGCTTT TTGAAACTGC
 651 CCAAACTGAG TTTCAAAGAT GCGGCGGTCA ACCGGGTGAT GAACAGAGAT
 701 GCGCCTCGgA TTTTgCGCGT GAgCTGGCGC CAGGTTTCCT TGGTATGCA
 751 CAGCATTTTC GCTGCTATG TGCATCGGG CAGCGTTTCA TGGATGATTT
 801 ACGCCGACCG CATGATGAGG CTCGCCACGG GCGTGTGGG GCGCGCAGCT
 851 GGTACGATT TGTGTCGCGAC TTGTGCAAA CATCGGCAA ACCAAGATAC
 901 GGAACAGTTT TCCGCCCTGC TCGACTGGGG TTTGCGCGTG TGCATGCTgc
 951 TGACGCTGCC GCGGgcGGTC GGACTGGCGG TGTTTGCTGT cCGcTGGGT
 1001 GCGACGCTGT TTATGTACCG CGwATTACG CTGTTTGACG CGCAGATGAC
 1051 GCACACGCGC CTGATTGCCT ATTCTTTCGG TTAATCGGC TTAATCATGA
 1101 TTAAGATGTT GGCACCCGGC TTCTATCGCG GGCAAAACAT CAwAmGCCC
 1151 GTCAAAATCG CCATCTTCAC GTCATCTCG mCGAGTTGA TGAACCTTGS
 1201 CTTTAYCGCG CCACTrAAC rCagTCGGAG TTTGCTGTGC CATCGGCTCG
 1251 GCGCGGTGTA TCAATGCGCG ATTGTGTGTT TACCTGTGTC GCAGACACGG
 1301 TATTATACCA CCTGG. CAAG GGTGGGCGAG GGTTCCT. AG CAAAAATGCT
 1351 GeTCTCGCTC CGCGTGA

This corresponds to the amino acid sequence <SEQ ID 114; ORF20>:

20 1 MNMLGALAKV GSLTMVSRVL GFVRDVTVIAR AFGAGMATDA FFVAFKLENI
 51 LRRVFAEAGAF AQAfVEILAE YKETSKEAA EFARHIVAGM LSFVLVIUTA
 101 LGILAAFWVVI YVSAPSFQAD ADKFQLSIDL LRITPFYILL ISLSSFVGSV
 151 LNSYHKGFIPI AFTPXFLMVS FIVFALFPVP YFDPVPVTAXA WAVFVGGILQ
 201 LXFQLPWLAK LGFLKLPLKS FKDAAVNRVM QKMAPAILGW SVAQSVLVIN
 25 251 TIFASYLQSG SVSWMYIADR MMELPSGVLG AALGTLILPT LSKHSAQNODT
 301 EQFSALLDWG LRLCMLITLP AAVGLAVLSF PLVATLFMYR XFTLFDAGMT
 351 QHALIAYSFG LIGLIMIKVL APGFYARQNI XXPVKIAIFT LICXQLMNLX
 401 FXGPLXIXGL SLAIGLGACI NAGLLFYLLR RHGIYQPKQG LGSVLXQKCC
 451 SRSP*

These sequences were elaborated, and the complete DNA sequence <SEQ ID 115> is:

30 1 ATGAATATGC TGGGAGCTTT GGCAAAAGTC GGCAGCCTGA CGATGGTGTG
 51 GCGCGTTTGG GGATTTGTGC GCGATACGGT CATTGCGCGG GCATTTGGCG
 101 CGGGTATGGC GAGCGATGCG TTTTGTGTGC CGTTCAAACG GCCCAACCTG
 151 CTTCGCCCGC TGTTTGGGGA GGGGCGCTTH GCGCAAGCGGT TTGTGCGCAT
 201 TTTGGCGGAA PACAAGGAAA CCGGTTCAA AGAGCGGCGG GAGGCTTTTA
 25 251 TCCGCCATGT GCGCGGAGTG CTGTGCTTGT TACTGCTTAT CGTTACCGCG
 301 CTGGGCATAC TTGCGCGCGC TTGGGTGATT TATGTTTCGG CACCCGGTGT
 351 TGCCCAAGAT GCGGACAAAT TTCAGCTCTC CATCGATTGT CTGCGGATTA
 401 CGTTTCCTTA TATATTATGT ATTTCCCTGT CTTCATTGTT CGGCTCGGTA
 451 CTCATTTCTT ATCATAAGTT CGGCAATCCG CGGTTTACGG CCACGTTTCT
 501 GAACGTGTGC TTTATCGTAT TCGCGCTGTT TTTGCTGCGC TATTTGCGATC
 551 GCGCCGTTCG CCGCTGGCGG TGGGCGGTCT TTGTGCGCGG CATTTTGCAA
 601 CTGCGGCTCC AACTGCCCTG GCTGGCGAAA CTGGGCTTTT TGAACATGCC
 651 CAAACTGAGT TTCAAAGATG CGGCGGTCAA CGCGCTGATT AACAGCATGG
 701 CGCCTGCGAT TTTGGGCGTG AGCGTGGCGC AGGTTTCTTT GGTGATCAAC
 751 ACGATTTCGC TGGCTTATCT GCAATCGGCG AGCGTTTCAT GGTGATTAAT
 801 CGCATCTTGC ATGATGAGAC TGCCCGCGCG CGTGTGGGG CGGCGATCTG
 851 GTACGATTTT GTCGCGGAGT TTGTCGCAAC ACTCGGCAAC CCAAGATGCA
 901 GAACAGTTTT CCGCCCTGCT CGAGTGGGCT TCGCGCTGCT GATGCTGCT
 951 GACGCTGCCG GCGGCGGTG GACTGGCGGT GTTGCTGCTC CGGCTGGTGG
 1001 CGACGCTGTT TATGTACCGG GAATTACCG TGTTTGACGC CAGATGAQC
 1051 CAACACGCGC TGATTGCCTA TTCTTTCGGT TTAATCGGCT TAATCATGAT
 1101 TAAAGTGTG GCACCCGGCT TCTATGCGCG GCAAAACATC AAAAAACCGC
 1151 TCAAAATCGC CATCTTCAGC CTCATCTGCA CGCAGTTGAT GAACCTTGCC
 1201 TTTATCGGCC CACTGAAACA CGTCGGACTT TCGCTTGCCA TCGGTCGGGG
 55 1251 CCGGTGTATC AATGCOGGAT TGTGTTTAT CCGTGTGGCG AGACACGGTA
 1301 TTTACCAACC TGCAAGGGT TGGGCAAGGT TCTAGGCAA AATGCTGCTC
 1351 TCGCTCGCG TGAATGCGCG CGGACTGTGG GCAGCGCAGG CTTACTCTGCC
 1401 GTTTGAATGG GCGGACGCGC GCGGAATGCG GAAGCGGGG CAGCTCTGCA
 1451 TCTTGATTGC CGTGCGCGCG GACTGTGATT TCGATCATCT GCGCGCTTTG
 60 1501 GGCTTCGCTG CCGGCCATTT CAAACGCGTG GAAACATGA

This corresponds to the amino acid sequence <SEQ ID 116; ORF20-1>:

1 MNMLGALAKV GSLTMVSRVL GFVRDVTVIAR AFGAGMATDA FFVAFKLENI
 51 LRRVFAEAGAF AQAfVPILAE YKETSKEAA EFARHIVAGM LSFVLVIUTA

101 LGILAAPWVI YVSAAGFAQD ADKQLSIDL LRITFFYILL ISLSSFVGSV
 151 INSYSKFCPI APTPTFLNV FIVFALFFVP YEDPPVTALA WAVEVGGILQ
 201 LGQLPWLAK LGFLKPKLS FKDAAVNRVM KOMAPAILGV SVQVSLVIN
 251 TIFASYLQSG SVSWMYADR MMELPSGVLG AALGTILTLP LSKHSANQDT
 301 EQFSALLDWG LRLCMLTLP AAVGLAVLS FLVATLIFYR EFTLFDQMT
 351 QHALIAYSFG LIGLIMIKVL APGFYARQNI KTPVKIAIFT LICTQLMNL
 401 FIGPLKHVGL SLAIGLGACI NAGLLFYLLR RHGIYQPGKG WAAFLAKMLL
 451 SLAVMCGGLW AAQAYLFFEW AHAGGMKAG QLCILIAVGG GLYFASLAAL
 501 GFPRHFHFRV EN*

10 Computer analysis of this amino acid sequence gave the following results:

Homology with the MviN virulence factor of *S. typhimurium* (accession number P37169)

ORF20 and MviN proteins show 63% aa identity in 440aa overlap:

Orf20	1	MMNLGALAKVGSGLTMVSRVLG	FVRD	TVIARAFGAGMATDAFFVAFKLPNLLRRVFAEGAF	60
MviN	14	MNLLKSLAAVSMTMESRVLGFARDAIVARIFGAGMATDAFFVAFKLPNLLRRIFAEGAF	73		
Orf20	61	AQAFVPILAAYKETRSKEAKEAFIRHVAGMLSFVLVITV	TALGILAAFPWVIYVSAPSFQAD	120	
MviN	74	SQAFVPIAAYKSKQGEATRFVAVVYSGLLTALAAVTVVAGMLAAPPWIMVTA	PGFADT	133	
Orf20	121	ADKQLSIDLLRITFFYILLISLSSFVGSVLSN	YHKFGIPAFTXPLNVSFIVFALFFVP	180	
MviN	134	ADKF L+ LLRITFFYILLISL+S VG++LN++++F	IPAF P FLN+S I FALF P	193	
Orf20	181	YEDPPVTAXAWAVFVGGILQLQFLPWLAKLGFLKPKLSFKDAAVNRVMKOMAPAILGV	240		
MviN	194	YFNPPVLAIAWAVTVGGVLQVYQLFYLKKTGMLVLPINFRD	GAMRVYQMGPAITLV	253	
Orf20	241	SVAQVSLVINTIFASYLQSGSVSWMYADRMELPSGVLG	AALGTILLPTLSKHSANQDT	300	
MviN	254	SVS+QSL+INTIFAS+L SSVSWMYADR+ME PSGLV	ALGTILLP+LSK A+ +	313	
Orf20	301	EQFSALLDWGLRLCMLTLPAAVGLAVLSFLVATL	IFYRKTFLFDAQMTQHALIAYSFG	360	
MviN	314	DEYCRMLDWGLRCLFLPALPSAVALGILAKPLTVSL	FGYKGFATFDAAMTQALIAYSVG	373	
Orf20	361	LIGLIMIKVLAPGFYARQNIIXPVKIAIFTLICXQLMNL	FXXXXXXXXXXXXXXXXXXXC	420	
MviN	374	LIGLIVVVLAPGFYSRQDIKTPVKIAIVTILMTQLMNL	AFIGPLKHAGLSLSIGLAACL	433	
Orf20	421	NAGLLFYLLRRHGIYQPGQ	440		
MviN	434	NASLLYQLRKQNIPTPQPG	453		

45 Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF20 shows 93.5% identity over a 447aa overlap with an ORF (ORF20a) from strain A of *N. meningitidis*:

orf20.pep	10	20	30	40	50	60
orf20a	10	20	30	40	50	60
orf20.pep	70	80	90	100	110	120
orf20a	70	80	90	100	110	120

	130	140	150	160	170	180
orf20.pep	ADK	FQ	L	S	I	D
orf20a	ADK	FQ	L	S	I	D
	130	140	150	160	170	180
orf20.pep	Y	F	D	P	P	T
orf20a	Y	F	D	P	P	T
	190	200	210	220	230	240
orf20.pep	F	D	P	P	T	A
orf20a	F	D	P	P	T	A
	190	200	210	220	230	240
orf20.pep	S	V	A	Q	V	S
orf20a	S	V	A	Q	V	S
	250	260	270	280	290	300
orf20.pep	S	V	A	Q	V	S
orf20a	S	V	A	Q	V	S
	250	260	270	280	290	300
orf20.pep	E	Q	F	S	A	L
orf20a	E	Q	F	S	A	L
	310	320	330	340	350	360
orf20.pep	E	Q	F	S	A	L
orf20a	E	Q	F	S	A	L
	310	320	330	340	350	360
orf20.pep	L	I	G	L	I	M
orf20a	L	I	G	L	I	M
	370	380	390	400	410	420
orf20.pep	N	A	G	L	F	Y
orf20a	N	A	G	L	F	Y
	430	440	450	460	470	480
orf20.pep	N	A	G	L	F	Y
orf20a	N	A	G	L	F	Y

The complete length ORF20a nucleotide sequence <SEQ ID 117> is:

1	ATGAATATGC	TGGGAGCTTT	GGTAAAAGTC	GGCAGCCTGA	CGATGGTGTC
51	GC	CGCGTTTTC	GGATTTGTGC	CGCATACGGT	CATTGCGGCG
101	CAGGCATGGC	GACGGATGCG	TTCTTTGTGC	CGTTCAAAC	GCCCAACCTG
151	CTTCGCCGCG	TGTTTGC	GGAAGGGGCGCTT	GCCCAAGCGT	TTGTGCCGAT
201	TTTGGCGGAA	TATAAGGAA	CGCGTCTTAA	AGAGCGCAGC	GAGGCTTTTA
251	TCCGCCATGT	GCGCGGGATG	CTGTGCTTTG	TACTGGTCA	CGTTACCGCG
301	CTGGGCATAC	TTGCGCGCGC	TTGGGTGATT	TATGTTTCCG	CACCGCGTTT
351	TGCCAAGAT	GCGCACAAAT	TTCAAGCTCT	TATCGATTG	CTCGCGATT
401	CGTTTCCTTA	TATCTTAATG	ATTTCACCTT	CCCTTTTGT	CGGCTCGGTA
451	CTCAATTCCT	ATCATAAAT	CAGCATTCCT	CGCTTTACG	CCAGCTTCCT
501	GAACGCTGCG	TTTATCGTAT	TCGCGCTGTT	TTTCGCGCG	TATTCGATTC
551	CTCCGCTTAC	CGCGCTGCT	TGGCGCGCTT	TTGCGCGCG	CATTTTGGA
601	CTCGGCTTCC	AACTGCCCTG	GCTGGCGAA	CTGGGTTT	TGAAACTGCC
651	CAAACTGAGT	TTCAAAGATG	CGCGCGTCAA	CCGCTGATG	AAACAGATGG
701	CGCGTGCAT	TTTGGCGGTG	AGCGTGGCG	AGATTCTTT	GGTGATCAAC
751	ACGATTTTTC	CGTCTTATCT	CGAATCGGCG	AGCGTTTCA	GGATGATTA
801	CGCGGACCGC	ATGATGGAA	TGCGCGCGCG	CGTGCTGGG	GCGGCACTCG
851	GTACGATTTT	GCTGCGGA	TTGTCAAAC	ACTCGGCA	CAAAGATACG
901	GAACGATTTT	CGCGCTGCT	CGACTGGGCT	TTGCGCTGTT	GCATGCTGCT
951	GACGCTGCGC	GCGCGCGTGC	GAATGCGGCT	GTTGCTGCT	CCGCTGGTGG
1001	CAACCTTGTT	TATGTACCGA	GAATTCACGC	TGTTTGACGC	GCAGATGACG
1051	CAACACGCGC	TGATTGCGTA	TTCTTTGCTG	TTAATCGGTT	TAATCATGAT
1101	TAAAGTGTG	GCGCGCGGCT	TTTATGCGCG	GCAAAATCAT	AAAACGCGCG
1151	TCAAAATCGC	CATCTTCAAG	CTCATTTCGA	CGAGTTGAT	GAACTTGGC
1201	TTTATCGCGC	CATCGAAACA	CGTGGGACTT	TCGCTTGCA	TCGCTTGGG
1251	CGCGCTTATC	AATCGCGGAT	TGCGCGGCT	CTGCTGCGC	AGACAGGTA
1301	TTTACCAAC	TGGCAAGGCT	TGGCGAGGCT	TTTGGCA	AATGCTGCTC
1351	TCGCTCGCGC	TGATGGGAGG	CGCGCTGAT	GCGGCCCAA	TCGCTGCTC
1401	CTTGCACTG	GCACACGCGC	CGGGAATGCA	AAGCGCGCGC	CGGCTCTTCA
1451	TCCTGATTGC	CGTGGCGGCG	GGACTGATT	TGCGATCACT	GCGCGCTTTC
1501	GGCTTCGCT	CGCGCATTT	CAACGCGGTG	GAAAGCTGA	

This encodes a protein having amino acid sequence <SEQ ID 118>:

```

1  MNMLGALVKV  GSLTMVSRVL  GFVRDVTIAR  AFGAGMATDA  FFVAFKLPNL
51  LRRVFAEGAF  AQAFVPILAE  YKETRSKEAT  EAFIRHVAGM  LSFVLVIUTA
101 LGILAAFPWI  YVSAPGFAKD  ADKFQSLIDL  LRITPPYILL  ISLSSFFGVS
151 LNSYHKFSIP  AFTPTFFLNVS  FIVFALFFVP  YFDPPTALA  WAFFVGGLQ
201 LGFQLPWLA  LGFLKLPKLS  FKDAAVNRVM  QMAPAILGV  SVAQISLVIN
251 TIFASYLQSG  SVSWMYYADR  MMELPGGVLC  AALGTILLPT  LSKHNSAQDT
301 EQFSALLDWG  LRXCMLLTLP  AAVGMAVLSF  PLVATLEWY  EFTLPDAQMT
351 QHALIAYSFG  LIGLIMIKVL  RAGFYARQNI  KTFVKIAIFT  LICTQLMNL
401 FIGPLKHVGL  LIGLIMIKVL  RAGFYARQNI  KTFVKIAIFT  LICTQLMNL
451 SLAVMGGGLY  RAQITWLPDW  AHAGGMQKAA  RLFIILAVGG  GLYFASLAA
501 GFRPRHFKRV  ES*

```

ORF20a and ORF20-1 show 96.5% identity in 512 aa overlap:

```

15  orf20a.pep      10      20      30      40      50      60
      MNMLGALVKVGSGLTMVSRVLGFVRDVTIARAFGAGMATDAFFVAFKLPNLLRRVFAEGAF
      orf20-1      MNMLGALAKVGSGLTMVSRVLGFVRDVTIARAFGAGMATDAFFVAFKLPNLLRRVFAEGAF
      10      20      30      40      50      60

20  orf20a.pep      70      80      90      100     110     120
      AQAFVPILAEYKETRSKEATEAFIRHVAGMLS FVLVIVITALGILAAFPWIIYVSAPGFAKD
      orf20-1      AQAFVPILAEYKETRSKEATEAFIRHVAGMLS FVLVIVITALGILAAFPWIIYVSAPGFAKD
      70      80      90      100     110     120

25  orf20a.pep      130     140     150     160     170     180
      ADKFQSLIDLLRITPPYILLISLSSFFGVS LNSYHKFSIP AFTPTFFLNVS FIVFALFFVP
      orf20-1      ADKFQSLIDLLRITPPYILLISLSSFFGVS LNSYHKFSIP AFTPTFFLNVS FIVFALFFVP
      130     140     150     160     170     180

30  orf20a.pep      190     200     210     220     230     240
      YFDPPTALAWAVFVGGLQLGFLPWLA LKLGFLKLPKLSFKDAAVNRVMQMAPAILGV
      orf20-1      YFDPPTALAWAVFVGGLQLGFLPWLA LKLGFLKLPKLSFKDAAVNRVMQMAPAILGV
      190     200     210     220     230     240

35  orf20a.pep      250     260     270     280     290     300
      SVAQISLVINTIFASYLQSGSVSWMYYADRM MELPGGVLC AALGTILLPT LSKHNSAQDT
      orf20-1      SVAQISLVINTIFASYLQSGSVSWMYYADRM MELPGGVLC AALGTILLPT LSKHNSAQDT
      250     260     270     280     290     300

40  orf20a.pep      310     320     330     340     350     360
      EQFSALLDWGLRXCMLLTLPAAVGMVLSFPLVATLFMYREFTLPDAQMTQHALIAYSFG
      orf20-1      EQFSALLDWGLRXCMLLTLPAAVGMVLSFPLVATLFMYREFTLPDAQMTQHALIAYSFG
      310     320     330     340     350     360

50  orf20a.pep      370     380     390     400     410     420
      LIGLIMIKVLAPGFYARQNIKTFVKIAIFT LICTQLMNLAFI GPLKHVGLSLAIGLACI
      orf20-1      LIGLIMIKVLAPGFYARQNIKTFVKIAIFT LICTQLMNLAFI GPLKHVGLSLAIGLACI
      370     380     390     400     410     420

55  orf20a.pep      430     440     450     460     470     480
      NAGLLFYLLRRHGIYQPGKWA AAFKAMLLSLAVMGGGLYAAQIWL PFDWAHAGGMQKAA
      orf20-1      NAGLLFYLLRRHGIYQPGKWA AAFKAMLLSLAVMGGGLYAAQIWL PFDWAHAGGMQKAA
      430     440     450     460     470     480

60  orf20a.pep      490     500     510
      RLFIILAVGGGLYFASLAAALGFRPRHFKRVESX
      orf20-1      QLCILIAVGGGLYFASLAAALGFRPRHFKRVENX
      490     500     510

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF20 shows 92.1% identity over a 454aa overlap with a predicted ORF (ORF20ng) from *N.*

gonorrhoeae:

5	orf20.pep	MMNLGALAKVGSILTMVSRVLGFVRDVTIARAFAGMATDAFFVAFKLPNLLRRVFAEGAF	60
	orf20ng	MMNLGALAKVGSILTMVSRVLGFVRDVTIARAFAGMATDAFFVAFKLPNLLRRVFAEGAF	60
10	orf20.pep	AQAFVPILAAYKETRSKEAEAFIRHVAGMLSFVLIVVTALGILAAPWVIYVSAPSAQD	120
	orf20ng	AQAFVPILAAYKETRSKEAEAFIRHVAGMLSFVLIVVTALGILAAPWVIYVSAPGFTKD	120
	orf20.pep	ADKFQLSIDLRLITFPFYILLISLSSFVGSVLNSYHKFGIPAPTPKFINVSPFIVFALFVFP	180
15	orf20ng	ADKFQLSISLRLITFPFYILLISLSSFVGSILNSYHKFGIPAPTTPTFLINSPFIVFALFVFP	180
	orf20.pep	YDPPVITAXAVFVGGILQLXQPLWLAKEFLKLPKLSFKDAVNVRVMQMAPAILGV	240
	orf20ng	YDPPVITAXAVFVGGILQLXQPLWLAKEFLKLPKLNFKDAVNVRVMQMAPAILGV	240
20	orf20.pep	SVAQSVLINTIFASYLQSGSVSWMYADRMMLPGSVLGAALGTILLPTLSKHSANQDT	300
	orf20ng	SVAQSVLINTIFASYLQSGSVSWMYADRMMLPGSVLGAALGTILLPTLSKHSANQDT	300
25	orf20.pep	EQFSALLDWGLRLCMLLTPAAVGLAVLSFPLVATLFMYRXTFLFDAQMTQHALIAYSFG	360
	orf20ng	EQFSALLDWGLRLCMLLTPAAVGLAVLSFPLVATLFMYRXTFLFDAQMTQHALIAYSFG	360
30	orf20.pep	LIGLIMIKVLAPGFYARQNIKXPKVIAIFTLICXQLMNLKFXGKPLXICGLSLAIGLGACI	420
	orf20ng	LIGLIMIKVLASGFYARQNIKTPVKIAIFTLICTQMLNLAIFGKPLKXAGLSLAIGLGACI	420
	orf20.pep	NAGLLFYLLRRHGIYQPKQLGSLVXQKCCSRSP	454
35	orf20ng	NAGLLFFLFRKHGIYRPGQGLGQPSWRKCCSRSP	454

An ORF20ng nucleotide sequence <SEQ ID 119> was predicted to encode a protein having amino acid sequence <SEQ ID 120>:

40	1	MMNLGALAKV	GSLTMVSRVL	GFVRDVTIAR	AFGAGMATDA	FFVAFKLPNL
	51	LRRVFAEGAF	AQAFVPILAE	YKETRSKEAT	EAFIRHVAGM	LSFVLIVVTA
	101	LGILAAPWVI	YVSAPGFTKD	ADKFQLSISL	LRITFPFYILL	ISLSSFVSGI
	151	LNSYHKFGIP	AETPTFLNIS	FIVFALFVFP	YDPPVITALA	WAVFVGGILQ
	201	LGQFLPWLAKE	FLKLPKLN	FKDAVNVRVM	QMAPAILGV	SVAQISLVIN
	251	TIFASYLQSG	SVSWMYADRM	MLPGSVLGA	ALGTILLPTL	LSKHSANQDT
45	301	EQFSALLDWG	LRLCMLLTP	AAAGLAVLSF	PLVATLFMYR	EFTLFDRQMT
	351	QHALIAYSFG	LIGLIMIKVL	ASGFYARQNI	KTVPKIAIFT	LICTQMLNLA
	401	FTGPKHAGL	SLATGLGACI	NAGLLFFLFR	KHGIYRPGQG	LQGPSWRKCC
	451	SRSP*				

Further DNA sequence analysis revealed the following DNA sequence <SEQ ID 121>:

50	1	ATGAATATGC	TTGGAGCTTT	GGCAAAGTC	GGCAGCCTGA	CGATGGTGTC
	51	CGCGGTTTTC	GGATTGTGTC	CGGATACGGT	CATTGCGCGG	GCATTGCGCG
	101	CGGGTATGCG	GACGATCGC	TTTTTTGTCG	CGTTCAAACT	GCCCAACTCG
	151	CTTGCGCGCG	TGTTTTCGGA	GGGGCGCTTT	GCCCAAGCGT	TGTGCGCCAT
	201	TTTGCGCGGAA	TATAAGGAAA	CGCGTTCTAA	AGAGGCGAcg	gAGGCTTTTA
55	251	TCCGCCACGT	tgccgggAatg	CTGTCGTTTG	TGCTGATcgt	cGttacCGCG
	301	CTGGGCATAC	TGCGCGCgcc	tTGGGTGATT	TATGTTtccg	CgcccGGCTT
	351	TACCAAGAC	GCGGACAACT	TCCAACCTTC	CATCAGCCTG	CTGCGGATTA
	401	CGTTTCTCTA	TATATTATGT	ATTTCITTTG	CTTCTTTTGT	CGGCTCGATA
	451	CTCAATCTCT	ACCATAGGTT	CGCATTCGCC	CGGTCATAGC	CCAGCTTTTC
	501	AAGCATCTCT	TTATCTCGAT	TCGCACTGTC	TTTCGCGCGG	TATTTTGCAC
60	551	CGCCCGTTAC	CGCGCTGCGC	TGGCGCGTTT	TTGTCGCGCG	TATTTTGCAC

601 CTCGGTTTCC AACTGCCGTG GCTGGCGAAA CTGGGCTTTT TGAACATGCC
 651 CAAACTGAAT TTCAAGATG CGCGCGTCAA CGCGCTCATG AAACAGATGG
 701 CGCTGCGAT TTTGGCGGTG agcgTGGCGC AAATTTCITT GgttatCAAC
 751 ACGATTTTTCG CgtCTTATCT GCAATCGGGC AGCGTTTCAT GGATGtatta
 801 cgCGACCCG ATGATGGAGc tgcgccGGGG CGTGTGGGG GCTGCATCG
 851 GTACAAATTTT GCTGCCGACT TTGTCAAAC ACTCGGCAAA CCAAGATACG
 901 GAACAGTTTT CCGCCTGCT CGACTGGGGT TTGCGCTGT GCATGCTGCT
 951 GACGCTGCCG CGCGCGGcg GACTGGCGGT ATTGCTGTT CCGCTGTTGG
 1001 CGAGCTGTGT TATGTACCGA GAATTCACGC TGTTTGACGC ACAAATGACG
 1051 CAACACGCGC TGATTGCCCTA TTCTTTCGTT TTAATCGGTT TAATTATGAT
 1101 TAAAGTTTG GCATCCGGCT TTTATGCGCG GCAAAACATC AAAACGCCCG
 1151 TCAAAATCGC CATCTTCACG CTCATCTGCA CGCAGTTGAT GAACCTCGCG
 1201 TTTATCGGTC CGTGAACA CGCGGGGCTT TCGCTCGCCA TCGGCTGGG
 1251 CGCGTGCATC AAGCGCGGAT TGTGTTCCTT CCTGTTCGCG AAACACGGTA
 1301 TTTACGGGCC cggcaggggt tgggcggcgt TCTTGGCGAA AATGCTGCTC
 1351 CGCGTCGCGC TGATGTCGCG CGGACTGTGG CGGCGCAGG CTTCCTCGCC
 1401 GTTCGAATGG GCGCACGCG CGCGAATGCG GAAAGCGGGG CAGCTCTGCA
 1451 TCGTGAATGC CGTCGCGCGG GACTGTATT TCGCATCTCT GCGCGCTTTG
 1501 GGCTTCGTC CGCGCCATT CAAACGCGTG GAAAGCTGA

20 This encodes the following amino acid sequence <SEQ ID 122; ORF20ng-1>:

1 MNMLGALAKV GSLTMVSRVL GFVRDTVIAR AFGAGMATDA FFVAFKLENL
 51 LRRVFPAEGAF AQAFVPIAE YKETSKEAT EAFIRHVAGM LSFVLIVVTA
 101 LGILAAFWVI YVSAFGFTKD ADKFLSISL LRITFPYILL ISLSSFVSGI
 151 LNSYHKFGIP AFTPTFLNIS FIVFALFFVP YFDPPTALA WAVEVSGTLQ
 201 LGFQPLWLAK LGFLKLPKLN FKDAAVNRVM QMAPAILGV SVAQISLVIN
 251 TIFASYLQSG SVSWMYADR MMELRRGVLS AALGTILLPT LSKHSANQDT
 301 EQFSALDWG LRLCMLLTLP AAAGLAVLSF PLVATLFMYR EFTLDAQMTQ
 351 QHALIAYSFG LIGLIMIKVL ASGFYARQNI KTVKILAIPT LICQILMNLA
 401 FIGPLKHAGL SLAIGLGACI NAGLFFLLR KHGIYRPGRG WAAFLAKMLL
 451 ALAVMCGGLW AAQAQLFFEW AHAGGMRKAG QLCILIAVGG GLYFASLAL
 501 GFRPRHFKRV ES*

ORF20ng-1 and ORF20-1 show 95.7% identity in 512 aa overlap:

		10	20	30	40	50	60
35	orf20-1.pep	MNMLGALAKV	GSLTMVSRVL	GFVRDTVIAR	AFGAGMATDA	FFVAFKLENL	RRVFPAEGAF
	orf20ng-1	MNMLGALAKV	GSLTMVSRVL	GFVRDTVIAR	AFGAGMATDA	FFVAFKLENL	RRVFPAEGAF
		10	20	30	40	50	60
40	orf20-1.pep	AQAFVPIAEY	KETSKEAAEA	FIRHVAGMLS	FVLIVTAL	LGILAAFWVI	YVSAFGFTKD
	orf20ng-1	AQAFVPIAEY	KETSKEATEA	FIRHVAGMLS	FVLIVTAL	LGILAAFWVI	YVSAFGFTKD
		70	80	90	100	110	120
45	orf20-1.pep	ADKFKLSIDL	LRITFPYILL	ISLSSFVSGI	SVNSYHKFGI	PAFTPTFLNIS	FIVFALFFVP
	orf20ng-1	ADKFKLSIDL	LRITFPYILL	ISLSSFVSGI	SVNSYHKFGI	PAFTPTFLNIS	FIVFALFFVP
		130	140	150	160	170	180
50	orf20-1.pep	YFDPPEVTAL	AWAVFVGGI	LQLGFQPL	WLAKLGFLK	PKLNFKDA	AVNRVMKQMAPAILGV
	orf20ng-1	YFDPPEVTAL	AWAVFVGGI	LQLGFQPL	WLAKLGFLK	PKLNFKDA	AVNRVMKQMAPAILGV
		190	200	210	220	230	240
55	orf20-1.pep	SVAQVSLVINTI	FASVLSQGS	SVSWMYADR	MMELRRGVLS	AALGTILLPT	LSKHSANQDT
	orf20ng-1	SVAQVSLVINTI	FASVLSQGS	SVSWMYADR	MMELRRGVLS	AALGTILLPT	LSKHSANQDT
		250	260	270	280	290	300
60	orf20-1.pep	EQFSALLDWGL	LRLCMLLTLP	AAAGLAVLSF	PLVATLFMYR	EFTLDAQMTQ	HALIAYSFG
	orf20ng-1	EQFSALLDWGL	LRLCMLLTLP	AAAGLAVLSF	PLVATLFMYR	EFTLDAQMTQ	HALIAYSFG
		310	320	330	340	350	360
65	orf20-1.pep	EQFSALLDWGL	LRLCMLLTLP	AAAGLAVLSF	PLVATLFMYR	EFTLDAQMTQ	HALIAYSFG
	orf20ng-1	EQFSALLDWGL	LRLCMLLTLP	AAAGLAVLSF	PLVATLFMYR	EFTLDAQMTQ	HALIAYSFG

		310	320	330	340	350	360
		370	380	390	400	410	420
5	orf20-1.pep	LIGLIMIKVLAPGFYARQNIKTPVKIAIFTLICTQLMNLAFIGPLKHVGLSLAIGLGACI					
	orf20ng-1	LIGLIMIKVLASGFYARQNIKTPVKIAIFTLICTQLMNLAFIGPLKHAGLSLAIGLGACI					
		370	380	390	400	410	420
		430	440	450	460	470	480
10	orf20-1.pep	NAGLLFYLLRRHGIYPGKGWAAFLAKMLLSLAVMCGGLWAAQAYLPFFWAHAGGMRKAG					
	orf20ng-1	NAGLLFFLLRKHGIYRPGRGWAAFLAKMLLALAVMCGGLWAAQACLFFWAHAGGMRKAG					
		430	440	450	460	470	480
		490	500	510			
15	orf20-1.pep	QLCILIAVGGGLYFASLAALGFRPRHFKRVENX					
	orf20ng-1	QLCILIAVGGGLYFASLAALGFRPRHFKRVESX					
		490	500	510			

20 In addition, ORF20ng-1 shows significant homology with a virulence factor of *S.typhimurium*:

sp|P37169|MVIN_SALTY VIRULENCE FACTOR MVIN pir||s40271 mvn protein - Salmonella typhimurium gI|438252 (Z26133) mviB gene product [Salmonella typhimurium] gnl|PID|d1005521 (D25292) ORF2 [Salmonella typhimurium] length = 524
 Score = 1573 (750.1 bits), Expect = 1.1e-220, Sum P(2) = 1.1e-220
 Identities = 309/467 (66%), Positives = 368/467 (78%)

25 Query: 1 MNMLGALAKVGSILTMVSRVLGFVRDVIARAFAGMATDAFFVAFKLENLLRRVFAEGAF 60
 MN+L +LA V S+TM SRVLGF RD ++AR FGAGMATDAFFVAFKLENLLRR+FAEGAF
 30 Sbjet: 14 MNLLKSLAAVSSMTMFSRVLGFARDIAVARIFAGMATDAFFVAFKLENLLRRIFAEAGF 73

Query: 61 AQAFVPIIAEYKETSKEATEAFIRHVGMLSFVLIVTALGILAAPFWIYVAPGFTKD 120
 +QAFVPIIAEYK + +EAT F+ +V+G+L+ L VVT G+LAAPWII V+APGF
 35 Sbjet: 74 SQAFVPIIAEYKSKQGEAATRIFVAYVSGLLTLALAVVTVAGMLAAPFWIMVTPAGFADT 133

Query: 121 ADKFLSISLLRITFPYILLISLSFVSGIILSYHKGIFAPFTPTFLNISPIVFALFFVP 180
 ADKF L+ LLRITFPYILLISL S VG+ILN++++F IPAF PTFINIS I FALF P
 40 Sbjet: 134 ADKFALTTQLLRITFPYILLISLASLVGAINTWNRFIPAFPTPTFLNISMIGFALFAP 193

Query: 181 YFDPPVTALAWAVFVGGLQLGFLPWLAKLGLFLKPLKLNFKDAVNRVMKQMAPAILGV 240
 YF+PPV ALAWAV VGG+LQL +QLP+L K+G L LP++NF+D RV+KQM PAIILGV
 45 Sbjet: 194 YFNPPVVALAWAVTVGGVQLQVLYQLPYLKKIGMLVLPRIINFRDITGAMRVVQKMGPAIILGV 253

Query: 241 SVAQISLIVINTIFASYLQSGSVSWMYADRMMLERGLVGAALGTILLPTLSKHSANQDT 300
 SV+QISL+INTIFAS+L SSGSVSWMYADRM+ME GVLG ALGTILLP+LSK A+ +
 50 Sbjet: 254 SVSQISLIINTIFASFLASGSVSWMYADRMLEFFSGVLGVALGTILLPLSLSKSFASGNH 313

Query: 301 EGFSAALLDWGLRLCMLLTPAAGLAVLSFPLVATLFMYREFTLFDQAQMTQHALIAYSFG 360
 ++ L+DWGLRLC LL LP+A L +L+ PL +LF Y +ET FDA MTQ ALIAYS G
 55 Sbjet: 314 DEYCRIMDWGLRLCFLLLALPSAVALGILAKPLTVSLFQYKGFATFADAMTQRALIAYSFG 373

Query: 361 LIGLIMIKVLASGFYARQNIKTPVKIAIFTLICTQLMNLAFIGPLKHAGLSLAIGLGACI 420
 LIGLI++KVLG GFY+RQ+IKTPVKIAI TLI TQLMNLAFIGPLKHAGLSL+IGL AC+
 60 Sbjet: 374 LIGLIVVAVLAPGFYSRQDIKTPVKIAIVTLTQLMNLAFIGPLKHAGLSLSIGLAACL 433

Query: 421 NAGLLFFLLRKHGIYRPGRGWXXXXXXXXXXXXXVMCGGLWAAQACL 467
 NA LL++ LRK I+ P GW VM L+ +P
 65 Sbjet: 434 NASLLYQLRKQNIITFPQPGWMMFLMRLIISVLVMAAVLFGVLHIMP 480

Score = 70 (33.4 bits), Expect = 1.1e-220, Sum P(2) = 1.1e-220
 Identities = 14/41 (34%), Positives = 23/41 (56%)

Query: 469 EWAHAGGMRKAGQLCILIAVGGGLYFASLAALGFRPRHFKR 509
 EW+ + +L ++ G YFA+LA LGF+ + F R
 70 Sbjet: 481 EWSQGLMNRLLRLMAVVIAGIAYFAALAVLGFVKVEFVR 521

Based on this analysis, including the homology with a virulence factor from *S.typhimurium*, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 15

5 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 123>:

```

1   atGATTAAAA   TCAAAAAAGG   TCTAAACCTG   CCCATCGCGG   GCAGACCGGA
51  GCAAGCCGTT   TACGACGGCC   CGGCCaTTAC   CGAAGTCGCG   TTGCTTGGCG
101 AAGAATAATG   CGGTATGCGC   CCTCGATGTA   AAGTCAAGGA   AGGCGATGCC
10  151  GTCAAAAAAG   GCCAAGTGCT   GTTTGAAGAC   AAAAAGAATC   CGGCGTGGT
201 GTTTACTCGG   CCGGCTTCAG   GCAAAATCGC   CGGATTTCAG   CGTGGCGAAA
251 AGCGCGTACT   TCAGTCAGTC   GTGATTGCGG   TTGAALGCAA   CGACGAATC
301 GAGTTTGAAC   GCTACGACCC   TGAAGCGCTG   GCAAACTTAA   CGGCGAAGA
351 AGTGGCGCGC   AACCTGATCC   AATCGGTTG   GTGGACTGCG   CTGGCACCC
15  401 GTCGGTTCAG   CAAAATTCCT   GCCGTCGATG   CCGAGCCGTT   GGCCATCTTC
451 GTCATGCGGA   TGGACACCAA   TCCG..

```

This corresponds to the amino acid sequence <SEQ ID 124; ORF22>:

```

1   MIKIKKGLNL   PIAGRFEQAV   YDGPATTEVA   LLGEYAYGMR   PSMKVKEGDA
51  VKKGQVLFD   KKNPGVVFTA   PASGKIAAIH   RGEKRVLQSV   VIAVEGNDEI
101 EFERYAPEAL   ANLSGEEVRR   NLIQSGLWTA   LRTRPFSKIP   AVDAEPFAIF
20  151  VNAMDTNP..

```

Further work revealed the complete nucleotide sequence <SEQ ID 125>:

```

1   ATGATTAAAA   TCAAAAAAGG   TCTAAACCTG   CCCATCGCGG   GCAGACCGGA
51  GCAAGCCGTT   TACGACGGCC   CGGCCATTAC   CGAAGTCGCG   TTGCTTGGCG
101 AAGAATAATG   CGGTATGCGC   CCTCGATGTA   AAGTCAAGGA   AGGCGATGCC
25  151  GTCAAAAAAG   GCCAAGTGCT   GTTTGAAGAC   AAAAAGAATC   CGGCGTGGT
201 GTTTACTCGG   CCGGCTTCAG   GCAAAATCGC   CGGATTTCAG   CGTGGCGAAA
251 AGCGCGTACT   TCAGTCAGTC   GTGATTGCGG   TGAAGGCAAA   CGGCGAAGA
301 GAGTTTGAAC   GCTACGACCC   TGAAGCGCTG   GCAAACTTAA   CGGCGAAGA
351 AGTGGCGCGC   AACCTGATCC   AATCGGTTG   GTGGACTGCG   CTGGCACCC
30  401 GTCGGTTCAG   CAAAATTCCT   GCCGTCGATG   CCGAGCCGTT   GGCCATCTTC
451 GTCATGCGGA   TGGACACCAA   TCCGCTGGCT   GCGACCCCTA   CGGTCAATTAT
501 CAAAGAAGCC   GCGAGGATT   TCAAAACGCG   CCTGTTGGTA   TTGAGCCGTT
551 TGACCGAAGC   CAAAATCCAT   GTTTGTAAGG   CAGCTGGCGC   AGACGTGGCG
601 TCTGAAAATG   CTGCCAACAT   CGAAACACAT   GAATTCGCGG   GCCGCGATCC
35  651 TGCCGGTTTG   AGTGGCAGCG   ACATTCAATT   CATCGAGCCG   GTCGGCGGCA
701 ATAAAAACCGT   GTGGACCATC   AATTATCAAG   ATGTAAATTA   CATTTGGCGGT
751 TTGTTTGCAA   CAGGCGCTCT   GAACACCGAG   CGCGTGATTG   CCTAGGTGG
801 GTCTCAAGTC   AACAAACCGC   GCCTTTGCGC   TACGCTTTTG   GGTGCGAAG
40  851 TATCGCAAAAT   TACTGCGGCG   AANTGCTTG   ACACGAGACA   CGAGTGAAAT
901 TCCGGTTCCG   TATTGAACGC   CGCGATTACA   CAAGCGCGCG   AGGATTAATT
951 GGGACGCTAC   CACAAATCAGA   TTTCCGTTAT   CGAAGAAGGC   CGCAGCAAG
1001 AGCTGTTCCG   CTGGGTTGCG   CGCAGCGCG   ACAAATCTC   CATCACCGGT
1051 ACAACCCCTG   GCCATTTCCT   GAAAAACAAA   CTCTCAAGT   TCAACACAGC
1101 CGTCAACGGC   GCGACCGCG   CCTGGTGCC   GATTGGTACT   TACAGCGCGG
45  1151 TGATGCCCTT   GATATCCTG   CCCACCTGCG   TTTTGCGGCA   TTTAATCGTC
1201 GCGGATACCG   ACAGCGCGCA   GGCATTGGGT   TGCTTGAAT   TGGACGAGA
1251 AGACCTCGCT   TTGTGCGAGT   TCGTCTGCC   GGGCAAAATC   GAATACGGCG
1301 CGCTGTTGCG   CAAAGTGCTG   GAAACCATTG   AGAAGGAAGG   CTGA

```

This corresponds to the amino acid sequence <SEQ ID 126; ORF22-1>:

```

50  1   MIKIKKGLNL   PIAGRFEQAV   YDGPATTEVA   LLGEYAYGMR   PSMKVKEGDA
51  VKKGQVLFD   KKNPGVVFTA   PASGKIAAIH   RGEKRVLQSV   VIAVEGNDEI
101 EFERYAPEAL   ANLSGEEVRR   NLIQSGLWTA   LRTRPFSKIP   AVDAEPFAIF
151 VNAMDTNP..   ADPTVIEKA   AEDPFRGLV   LSLRTERKH   VCKAAGADVP
201 SENAANIETH   EFGGHPAAGL   SGTHTFIEP   VGANKTVWT   NYQDVITIGR
55  251  LFATGRINTE   RVIALGGSQV   NKFRLLRTVL   GAKVQTTAG   ELVDITDNRV

```

301 SGSVLNGAIT QGAHDYLGRI HNQISVIEEG RSKELFGWVA PQDPKYSITR
 351 TTLGHFLKNK LFKFTTAVNG GDRAMVPIGT YERVMPLDIL PTLLLRDLIV
 401 GDTSDAQLAG CLELDEEDLA LCSFVCPGKY EYGLLRVLV ETXKEG*

Further work identified the corresponding gene in strain A of *N.meningitidis* <SEQ ID 127>:

5 1 ATGATTAAAA TCAAAAAGG TCTAAACCTG CCGATCGCGG CGAGACCGGA
 51 GCAAGTCATT TATGACGGCG CCGTCACTAC CGAAGTCGCG TTGCTTGGCG
 101 AAGAATATGC CGGTATGCGC CCTTNGATGA AAGTCAAGGA AGCGGATGCC
 151 TCTAAAAGAG GCCAAGTGCT GTTTGAAGAC AAAAAGNATC CGGCGGTGGT
 201 GTTTACCGCG CCGTGTTCAG GCAAAATGCG CGCCATCCAT CGCGCGGAAA
 251 AGCGGCTACT TCAGTCGGTG GTGATTCGCG TTGAAGGCAA CGAOGAARTC
 301 GAGTTCGAAC GCTACGCGCC CGAAGCGITG CGCACTTAA CGCGCGGANGA
 351 ANTNNNGNC AATCTGATCC AATCCGGTTT GTGGACTGCG CTGCGTANCC
 401 GPCGCTTCAG CAAAATCCCT GCGCTCGATG CGGAGCCGTT GCGCATCTTC
 451 GTCAATGCGA TGGACACCAA TCCGCTNGCG CGAGACCCGT TGGTTGTGAT
 501 CAAAGAAGCC GNCGANGATT TCAGACGANG TNGCTGGTA TTGAGCCGTT
 551 TGACCGAGCG TAAAATCCAT GTGTGTAAGG CAGCTGGCGC AGAOGTGGCG
 601 TCTGAAATATG CTGCCAAACAT CGAAACACAT GAATTGGGCG GCGCGCATCC
 651 GCGCGTTTGG AGTGCCACGCG ACATTCAATT CATTTAGCGG CTGCGTGCAG
 701 ACAGAACCGT TPGGACCATC ATTATCAAGC CATCGAGCGT CATCGAGCGT
 20 751 TGTCTTGCAA CAGCCGCTCT GAACACCGAG CCGCTGATTG CTTTGGTGTT
 801 TTCTCAAGTC AACAAACAC GCCTCTTGCG TACCGTTTGG GGTGCGAAG
 851 TATGCGAAAT TACTGCGGCG GAATTGGTTG CAGCAGACAA CGCGCTGATT
 901 TCCGGTTCG TATTGAACGG CGCGATTACA CAGGGCGCGC ACGATTATT
 951 GGGACGCTAC CACAATCAGA TTTCGGTTAT CAAAGAGCG CGCAGCAGAG
 25 1001 AGCTGTTGCG CTGGGTTGCG CGCGAGCGCG CAAATATCTC CATCAGCGGT
 1051 ACGACCCCTG GCCATTTCCT GAAAAACAAA CTCTTCAAGT TCACGACGAC
 1101 GCTCAACGGT GGCACCGCG CCAATGTTGCC GATTGGTACT TACGAGCGCG
 1151 TAATGCCGCT AGACATCTGT CTAACCTGCG TTTTGGCGGA TTTAATCGCT
 1201 GGCGATACCG ACAGCGCGCA AGCAATTGGT TCGTTGGAAT TGGACGAAGA
 30 1251 AGACCTCGCT TTGTGCGAGT TCGTCTGCCG GGGCAGAAAT GAATTANGGCC
 1301 CGCTGTTGCG TAAGGTGCTG GAAACNTTGG AGAAGAAGG CTGA

This encodes a protein having amino acid sequence <SEQ ID 128; ORF22a>:

35 1 MIKIKKGLNL PIAGRPEQVI YDGPVITEVA LLGEEYAGMR PMXKVKEGDA
 51 VKKGQVLFED KKXPGVFTTA PVSGKIAIHH RGEKRVLSQV VIAVEGNDIEI
 101 EFERYAPEAL ANLSGXEXXX NLIQSLWTA LRKPPFSKIP AVDAEPPAIF
 151 VNAMDTNPLA ADPVVVIKEA XXDFRRXXLV LSRITFERKH VCKAAGADVP
 201 SENAANIETH EFGGPHAPGL SGTTHIFTEP VGANKTVMTI NYQDVIAIGR
 251 LEFATGRNTE RVIALGSSQV NKPELLRTVL GAKVSIQTAG ELVDADNRVI
 40 301 SGSVLNGAIT QGAHDYLGRI HNQISVIEEG RSKELFGWVA PQDPKYSITR
 351 TTLGHFLKNK LFKFTTAVNG GDRAMVPIGT YERVMPLDIL PTLLLRDLIV
 401 GDTSDAQLAG CLELDEEDLA LCSFVCPGKY EXGELLRLVL ETXKEG*

The originally-identified partial strain B sequence (ORF22) shows 94.2% identity over a 158aa overlap with ORF22a:

		10	20	30	40	50	60
45	orf22.pep	MIKIKKGLNLPIAGRPEQVYDGPVITEVALLGEEYAGMRPSMKVKKGDAVKKGQVLFED					
	orf22a	MTKIKKGLNLPIAGRPEQVYDGPVITEVALLGEEYAGMRPSMKVKKGDAVKKGQVLFED					
		10	20	30	40	50	60
50	orf22.pep	KKNPGVVFTAPASGKIAIHRGEKRVLSQVVIIVAVENDEIEFERYAPEALANLSGEEVRR					
	orf22a	KKKPGVVFTAPVSGKIAIHRGEKRVLSQVVIIVAVENDEIEFERYAPEALANLSGXEXXX					
		70	80	90	100	110	120
55	orf22.pep	NLIQSLWTALRTRPFSKIPAVDAEPPAIFVNAMDTNP					
	orf22a	NLIQSLWTALRTRPFSKIPAVDAEPPAIFVNAMDTNPLAADPVVVIKEAXXDFRRXXLV					
60		130	140	150	160	170	180

The complete strain B sequence (ORF22-1) and ORF22a show 94.9% identity in 447 aa overlap:

		10	20	30	40	50	60
	orf22a.pep	MIKIKKGLNLPIAGRPEQVIYDGPVITEVALLGEEYAGMFPXMKVKEGDAVKKGQVLFED					
5	orf22-1	MIKIKKGLNLPIAGRPEQAVYDGPVITEVALLGEEYAGMFPXMKVKEGDAVKKGQVLFED					
		10	20	30	40	50	60
	orf22a.pep	KKXPGVVFAPVSGKIAAIHRGEKRVLQSVVIAVEGNDEIEFERYAPEALANLSGKEXXX					
10	orf22-1	KKNPGVVFAPVSGKIAAIHRGEKRVLQSVVIAVEGNDEIEFERYAPEALANLSGEEVRR					
		70	80	90	100	110	120
	orf22a.pep	NLIQSGLWLTALRXRPFISKIPAVDAEPFAIFVNAMDTNPLAADPVVVIKEAXXDFRXXLV					
15	orf22-1	NLIQSGLWLTALRXRPFISKIPAVDAEPFAIFVNAMDTNPLAADPTVVIKEAADFKRGLLV					
		130	140	150	160	170	180
	orf22a.pep	LSRLTERKIHVCKAAGADVSENAAIETHFEFGGPHAGLSGTHIHFIIEPVGANKTVMVTI					
20	orf22-1	LSRLTERKIHVCKAAGADVSENAAIETHFEFGGPHAGLSGTHIHFIIEPVGANKTVMVTI					
		190	200	210	220	230	240
	orf22a.pep	NYQDVIAIGRLFATGRINTERVIALGGSQVKNPRLRLTVLGAKVSQLTAGELVDADNRVI					
25	orf22-1	NYQDVITIGRLFATGRINTERVIALGGSQVKNPRLRLTVLGAKVSQLTAGELVDTDNRVI					
		250	260	270	280	290	300
	orf22a.pep	SGSVLNGAITQGAHDYLGRYHNQISVIEEGRSKELFGWVAPQPDKYSITRTTLGHFLKNN					
30	orf22-1	SGSVLNGAITQGAHDYLGRYHNQISVIEEGRSKELFGWVAPQPDKYSITRTTLGHFLKNN					
		310	320	330	340	350	360
	orf22a.pep	LKFFTTAVNGGDRAMVPITGYERVMPDLITPLLRLDLIVGDTSAQALGCELEDEEDLA					
35	orf22-1	LKFFTTAVNGGDRAMVPITGYERVMPDLITPLLRLDLIVGDTSAQALGCELEDEEDLA					
		370	380	390	400	410	420
	orf22a.pep	LCSFVCPGKYEXGPELLRKVLETKEGEX					
40	orf22-1	LCSFVCPGKYEXGPELLRKVLETKEGEX					
		430	440				
45	orf22a.pep	LCSFVCPGKYEXGPELLRKVLETKEGEX					
	orf22-1	LCSFVCPGKYEXGPELLRKVLETKEGEX					

Further work identified a partial gene sequence <SEQ ID 129> from *N.gonorrhoeae*, which encodes the following amino acid sequence <SEQ ID 130; ORF22ng>:

50	1	MIKIKKGLNLPIAGRPEQVIYDGPVITEVALLGEEYAGMFPXMKVKEGDAVKKGQVLFED
	51	VKKGQVLFEDKKNPGVVFAPVSGKIAAIHRGEKRVLQSVVIAVEGNDEIEFERYAPEALANLSGKEXXX
	101	EFERYVPEALAKLSSEKVRRLNLIQSGLWLTALRXRPFISKIPAVDAEPFAIFVNAMDTNPLAADPVVVIKEAXXDFRXXLV
	151	VNAMDTNPLAADPTVVIKEAADFKRGLLVLSRLTERKIHVCKAAGADVSENAAIETHFEFGGPHAGLSGTHIHFIIEPVGANKTVMVTI
	201	SENAAIETHFEFGGPHAGLSGTHIHFIIEPVGANKTVMVTI
55	251	LFVTGRNLNTE RVVALGGLQV NKPRLRLTVL GAKVSQLTAG ELVDADNRVI
	301	SGSVLNGAIA QGAHDYLGRY HN*

Further work identified complete gonococcal gene <SEQ ID 131>:

60	1	ATGATTAAATCAAAAAAGGCTCTAAATCTGCCATCGCGCGACACCGGA
	51	GCAAGTCATTATGACGCGCCGGCCATTACGAAAGTCGCGTTGCTTGCGG
	101	AAGATATGTCGGCATCGCCCTCGATGATAAATCAAGGAAGTGGAAGCC
	151	GTCAAAAAGGCCAAGTGCTGTTTGAAGCAAAAAGATCGGGCGTAGT
	201	ATTACTGCGCGGCTTCAGCAAAATCGCGCTATTACCGTGGCGAAA
	251	AGCGCGTACTTCAGTCAGCTGTGATTGCGGTGAAGGCCAACGACGAATC
	301	GAGTTCGACGCTACGTACGTGAAGCGCTGCATAATGACGACGAAAA

5
10
15
20

```

351 AGTGC GCGC AACCTGATC AATCAGGCTT ATGGA CTGGC CTTGCGCACCC
401 GTCCGTTT CAG CAAAATCCGT GCGGTAGATC CCGAGCCGTT GCGCATCTTC
451 GTCAATGCGA TGGACACCAA TCCGCTGGCT CCGCACCTTA CCGTATCAT
501 CAAAGAGCC GCGCAAGACT TCAACGCGG CCGTGTGGTA TTGAGCGCGC
551 TGACCAAGC TAAAATCCAT GTGTGAAGC CAGCAGGCGC AGACGTGCGG
601 TCTGAAATG CTGCCAATAT CGAACAACAT GAATTTGGG GCCCGCATCC
651 TGCCGCGTTG AGTGGCACGC ACATTATTT CATCGAGCCA GTCCGCGCGA
701 ATAAAAACCG GTGGACCATT AATTATCAAG ACGTGATTGC TATCGGACGT
751 TTGTTCGTAA CAGGCGCTCT GAATACCGAG CCGGTGGTTG CTTTGGCGGG
801 CCTGCAAGTC AACAAACCGC GCCTCTTGGC TACCGTTTGG GGTGCGAAGG
851 TGTCTCAACT TACCGCCGGC GAATTTGGTT CAGCGGACAA CCGCGTGATT
901 TCCGGTTCGG TATTGAACGG TCGGATTGCA CAAGGCGCGC ATGATTATTT
951 GGGACGCTAC CACAATCAGA TTTCGGTAT CGAAGAAGGC CGCAGCAAG
1001 AGCTGTTCCG CTGGGTTGCG CCGCAGCGCG ACAAATACTC CATCACGCGC
1051 ACCACTCTCG GCCATTTCCT AAAAAACAAA CTCCTCAAGT TCACGACACG
1101 CGTCAACGGC GCGACCCGGC CCAATGTTAC GATCGGCACT TATGAGCGCG
1151 TAATGCGGTT GGACATCCGT CTTACCTTGC TTTTGGCGGA TTTAATCGTC
1201 GCGGATACCG ACAGCGCGCA GCCTTGGGTT TCGTGGAGAT TGGACGAAGA
1251 AGACCTCGCT TTGTGACGCT TCGTCTGCC GGCACAAATC GAATACGCGC
1301 CGCTGTTGCG CAAAGTGCTG GAAACCATTG AGAAGAGGCT CTGA

```

This encodes a protein having amino acid sequence <SEQ ID 132; ORF22ng-1>:

25
30

```

1 MIKIKKGLNL PIAGRPEQVI YDGPATTEVA LLGEEYVGM RPSMKIKEGEA
51 VKKGQVLFED KKNPGVVFTA PASGKIAIHR GERKRVLSQV VIAVEGNDEI
101 EFERYVPEAL AKLSSEKVRN NLIQSGLWTA LRTTRPFSKI PAVDAEFPFAIF
151 VNAMDTNPLA ADPTVIEKEA AEDFKRLLV LSRLETKIH VCKAAGADVP
201 SENANINETH EFGGPHIAGL SGTHIFTEP VGANKTWTIT NYQDVIAIGR
251 LFVTLGRINTE RVVALGGLQV NKPRLLRTVL GAKVSLTAG SLVDADNRVI
301 SGSVLNGAIA QCAHDYLGRI HNQISVIEEG RSKELGWVA PDPDKYSITR
351 TTLGHFLRKN LFKFTTAVNG GDRAMVPIGT YERVMFLDIL PTLLLRDLIV
401 GDTBSAQALG CLELDEEDLA LCSFVCPGKY EYGPLLRKLV ETLEKRG*

```

The originally-identified partial strain B sequence (ORF22) shows 93.7% identity over a 158aa overlap with ORF22ng:

35
40
45

```

orf22.pep MIKIKKGLNLPIAGRPEQVIYDGPATTEVALLGEEYAGMRPSMKVKEGDAVKKGQVLFED 60
orf22ng MIKIKKGLNLPIAGRPEQVIYDGPATTEVALLGEEYVGM RPSMKIKEGEAVKKGQVLFED 60
orf22.pep KKNPGVVFTAPASGKIAIHRGERKRVLSQV VIAVEGNDEIEFERYAPEALANLSGEEVRR 120
orf22ng KKNPGVVFTAPASGKIAIHRGERKRVLSQV VIAVEGNDEIEFERYVPEALAKLSSEKVRN 120
orf22.pep NLIQSGLWLTALRTRPFSKI PAVDAEFPFAIFVNAMDTNP 158
orf22ng NLIQSGLWLTALRTRPFSKI PAVDAEFPFAIFVNAMDTNPLAADPTV IIEKAAEDFKRGLLV 180

```

The complete sequences from strain B (ORF22-1) and gonococcus (ORF22ng) show 96.2% identity in 447 aa overlap:

50
55
60

```

orf22-1.pep 10 20 30 40 50 60
MIKIKKGLNLPIAGRPEQVIYDGPATTEVALLGEEYAGMRPSMKVKEGDAVKKGQVLFED
orf22ng-1 10 20 30 40 50 60
MIKIKKGLNLPIAGRPEQVIYDGPATTEVALLGEEYVGM RPSMKIKEGEAVKKGQVLFED

orf22-1.pep 70 80 90 100 110 120
KKNPGVVFTAPASGKIAIHRGERKRVLSQV VIAVEGNDEIEFERYAPEALANLSGEEVRR
orf22ng-1 70 80 90 100 110 120
KKNPGVVFTAPASGKIAIHRGERKRVLSQV VIAVEGNDEIEFERYVPEALAKLSSEKVRN

orf22-1.pep 130 140 150 160 170 180
NLIQSGLWLTALRTRPFSKI PAVDAEFPFAIFVNAMDTNPLAADPTV IIEKAAEDFKRGLLV

```

orf22ng-1	NLIQSGLWALTRTPFSKIPAVDAEFPFAIVFNAMDTNPDAADPTVIKEAAEDFKGLLV	130	140	150	160	170	180
orf22-1.pep	LSRLTERKIHVCKAAGADVPSENAANIETHEFGGPHFAGLSGTHIFIEPVGANKTWTWT	190	200	210	220	230	240
orf22ng-1	LSRLTERKIHVCKAAGADVPSENAANIETHEFGGPHFAGLSGTHIFIEPVGANKTWTWT	190	200	210	220	230	240
orf22-1.pep	NYQDVTITGRLFATGRINTEPVIALGGSQVNKPRLLRTVLGAKVSGITAGELVDNDNRVI	250	260	270	280	290	300
orf22ng-1	NYQDVIAIGRLFVTGRINTEPVIALGGLOVNKPRLLRTVLGAKVSGITAGELVDADNRVI	250	260	270	280	290	300
orf22-1.pep	SGSVLNGAITQGAHDYILGRYHNQISVIEEGRSKELFGWVAQPDPKYSITRTTLGHPLKNN	310	320	330	340	350	360
orf22ng-1	SGSVLNGAIAQGAHDYILGRYHNQISVIEEGRSKELFGWVAQPDPKYSITRTTLGHPLKNN	310	320	330	340	350	360
orf22-1.pep	LFKFNTAVNGGDRAMVPIGYTERVMPDLIDITLLLRDLIVGDTSDAQLGCLLEDEEDA	370	380	390	400	410	420
orf22ng-1	LFKFTTAVNGGDRAMVPIGYTERVMPDLIDITLLLRDLIVGDTSDAQLGCLLEDEEDA	370	380	390	400	410	420
orf22-1.pep	LCSFVCPGKYEYGP LLRKVLLETIEKEGX	430	440				
orf22ng-1	LCSFVCPGKYEYGP LLRKVLLETIEKEGX	430	440				

Computer analysis of these sequences gave the following results:

Homology with 48kDa outer membrane protein of *Actinobacillus pleuropneumoniae* (accession number U24492).

ORF22 and this 48kDa protein show 72% aa identity in 158aa overlap:

Orf22	1	MIKIKKGLNLPDIAAGPEQVVDGPAITEVALLEGEEYAGMRSPMKVYEGDVAVKKQGVLFED	60
48kDa	1	MI IKKGL+LPDAG P Q + + + + + E + V + LGEEY GMRSPMKV + EGD VKKQGVLFED MI*TKKGLDLPDAGTAPQVIHNHGKVTNVEAMVLGEEYVGMRSMPKVREGDVVKKQGVLFED	60
Orf22	61	KKNPQGVVFTAPASGKIAI+HGEKRVQLQSVIVAXNDEIEFERYAPFALNLSGEEVRR	120
48kDa	61	KKNPQGVVFTAPASG + I + HGEKRVQLSVI VE + + + I F R Y LA+LS E + V + + + KKNPQGVVFTAPASGTVVTTNKIKGKRVQLQSVIVKVEGDEQITFTTRYEAAQLASLSAEQVVKQ	120
Orf22	121	NLIQSGLWTAALTRFPSPKIPADVAEFPFIVSNMNTNP	158
48kDa	121	NLI+SGLWTA TRFPSPK+PA+DA P + I + FVSNMNTNP	158
Orf22	121	NLIQSGLWTAFTTRFPSPKIPALDAPSSIFIVSNMNTNP	158

ORF22a also shows homology to the 48kDa *Actinobacillus pleuropneumoniae* protein:

gil1185395 (U24492) 48 kDa outer membrane protein [Actinobacillus pleuropneumoniae]
Length = 449

Score = 530 bits (1351), Expect = e-150
Identities = 274/450 (60%), Positives = 323/450 (70%), Gaps = 4/450 (0%)

```
Query: 1 MIKIKKGLNPIAGRPGQVIYDGPVITEVALLGEEYAGMRPMKMKVKGDAVKKQGVLFED 60
      1 MI IKKGL+LPIAG P QVI+H + EVA+GEEY GMRP MKV+EGD VKKGQGVLFED
Sbjct: 1 MITIKKGLDLPIGTAIPQAQVI+HNGTIVNEVAMIGEEYGMRPMSKVKVKGDAVKKQGVLFED 60

Query: 61 KKKPGVVFPTAPVSGKIAAIHRGKRVLSQVIAVEGNDRIEFERYAPEALNLSGXEXXX 120
      61 KK PGVVFPTAP SG + I+RGKRVLSQVI VEG+++I F RY LA+LS +
Sbjct: 61 KKNPGVVFPTAPAGSTVITINGRKGKRVLSQVIVKEGEGDITFFRYAEQAALSLSAEQVK 120
```

Query: 121 NLIQSGSLWTLALRXPFSKIPAVDAEPFAIFVNMADTNPLAADPVPVVVKEAXDXFRXXLV 180
 NLI+SGLWTA R RPFSK+PA+DA P +IFVNMADTNPLAADP VV+KE DF+ V
 Sbjct: 121 NLIQSGSLWTAFTTRPFSKVPALDAIPSSIFVNMADTNPLAADPEVVLKEYETDFDKGLTV 180

5 Query: 181 LSRL--TERKIHVCKAAGADVP-SENAANIETHEFGGPHAGLSGTHIHFIETPVGANKTV 237
 L+RL ++ +++CK A +++P S I F G HPAGL GTHIHFI++PVGA K V
 Sbjct: 181 LTRLFNGQKPVYLCKDADSNIPLSFAIEGITIKSFSGVHPAGLVGTHIHFIHFDVPVGATKQV 240

10 Query: 238 WTINYQDVIAIGRLFTVTRINTERVVALGGLOVKNPRLRLTVLGAKVSQLTAGELVDADN 297
 W +NYQDVIAIG+LF TG L T+R+I+L G QV PRL+RT LGA +SQT+LA EL +N
 Sbjct: 241 WHLNYQDVIAIGKLTFTGELFTDRITISLAGPQVKNPRLVTRLGANLSQLTANELNAGEN 300

15 Query: 298 RVISGSLVNGAITQGAHDYLGRYHNQISVIEEGRSKELFGWVAPQPDKYSITRRTLGHFL 357
 RVISGSLV+GA G DYLGRY Q+SV+ EGR KSLFGW+ P DK+SITRT LGHF
 Sbjct: 301 RVISGSLVSGATAAGPVDYLGRYALQVSVLAEGREKELFGWIMPBGDKFSITRRTLVLGHFG 360

20 Query: 358 KNKLFKFTTAVNGGDRAMVP+IGTYERVVMXXXXXXXXXXXXXVGTDSAQXXXXXXXXXX 417
 K KLF FTTAV+GG+RAMVPIG YERV GDTDSAQ
 Sbjct: 361 K-KLFFNTTAVHGERAMVPIGAYERVMPLDIPTLLRLDLAAGDTDSAQNLGCLDEE 419

Query: 418 XXXXSFVCPGKYEKGXGLLRKVLKTEKEG 447
 ++VCPGK GP+LR LE EKEG

ORF22ng-1 also shows homology with the OMP from *A. pleuropneumoniae*:

gi|1185395 (U24492) 48 kDa outer membrane protein [Actinobacillus
 pleuropneumoniae] Length = 449
 Score = 555 bits (1414), Expect = e-157
 Identities = 284/450 (63%), Positives = 337/450 (74%), Gaps = 4/450 (0%)

25 Query: 27 MIKIKKGLNLPIAGRPEQVIYDGPATEVALLGEEYVGMRRPSMKIKEGAVKKGQVLFED 86
 MIKIGL+LPIAG P QVI++G +EVA+LGEYVGMRRPSMK++EG+ VKKGQVLFED
 Sbjct: 1 MITIKKGLDLPIAGTPAQVHNGNTVNEVAMLGEEYVGMRRPSMKVREGVVKKGQVLFED 60

30 Query: 87 KKNPGVVFTAPASGKTIARIHGEKRVLSVIVAEVGNDEIFERYVPEALAKLSSEKVR 146
 KKNPGVVFTAPASG +I+RGEKRVLSVVI VEG+++I F RY LA LS+E+V++
 Sbjct: 61 KKNPGVVFTAPASGTVTTINRGEKRVLSVVIKVEGDEQITPTTRYEAQGLASAEQVQK 120

35 Query: 147 NLIQSGSLWTLALRXPFSKIPAVDAEPFAIFVNMADTNPLAADPTVILKEAEDFKRGLLV 206
 NLI+SGLWTA RTRPFSK+PA+DA P +IFVNMADTNPLAADP V++KE DFK GL V
 Sbjct: 121 NLIQSGSLWTAFTTRPFSKVPALDAIPSSIFVNMADTNPLAADPEVVLKEYETDFDKGLTV 180

40 Query: 207 LSRL--TERKIHVCKAAGADVP-SENAANIETHEFGGPHAGLSGTHIHFIETPVGANKTV 263
 L+RL ++ +++CK A +++P S I F G HPAGL GTHIHFI++PVGA K V
 Sbjct: 181 LTRLFNGQKPVYLCKDADSNIPLSFAIEGITIKSFSGVHPAGLVGTHIHFIHFDVPVGATKQV 240

45 Query: 264 WTINYQDVIAIGRLFTVTRINTERVVALGGLOVKNPRLRLTVLGAKVSQLTAGELVDADN 323
 W +NYQDVIAIG+LF TG L T+R+I+L G QV PRL+RT LGA +SQT+LA EL +N
 Sbjct: 241 WHLNYQDVIAIGKLTFTGELFTDRITISLAGPQVKNPRLVTRLGANLSQLTANELNAGEN 300

50 Query: 324 RVISGSLVNGAITQGAHDYLGRYHNQISVIEEGRSKELFGWVAPQPDKYSITRRTLGHFL 383
 RVISGSLV+GA A G DYLGRY Q+SV+ EGR KELFGW+ P DK+SITRT LGHF
 Sbjct: 301 RVISGSLVSGATAAGPVDYLGRYALQVSVLAEGREKELFGWIMPBGDKFSITRRTLVLGHFG 360

55 Query: 384 KNKLFKFTTAVNGGDRAMVP+IGTYERVVMXXXXXXXXXXXXXVGTDSAQXXXXXXXXXX 443
 K KLF FTTAV+GG+RAMVPIG YERV GDTDSAQ
 Sbjct: 361 K-KLFFNTTAVHGERAMVPIGAYERVMPLDIPTLLRLDLAAGDTDSAQNLGCLDEE 419

60 Query: 444 XXXXSFVCPGKYEGBGLLRKVLKTEKEG 473
 ++VCPGK YGP+LR LE TEKEG
 Sbjct: 420 DLALCTYVCPGKNNYGPMLRAALEKTEKEG 449

Based on this analysis, including the homology with the outer membrane protein of *Actinobacillus pleuropneumoniae*, it was predicted that these proteins from *N. meningitidis* and *N. gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF22-1 (35.4kDa) was cloned in pET and pGex vectors and expressed in *E. coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 5A shows the results of affinity purification of the GST-fusion protein, and Figure 5B shows the results of expression of the His-fusion in *E. coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result) and FACS analysis (Figure 5C). These experiments confirm that ORF22-1 is a surface-exposed protein, and that it is a useful immunogen.

Example 16

The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 133>:

```

1  ..GCGnCGnAAA TCATCCATCC CC..nACGTC GTAGGCCCTG AAGCCCAATG
51  GTTTTTTATG GTAGCCAGTA CGTTTGTGAT TGCITTTGATT GGTATTATTG
101  TTA CTGAAAA AATCGTCGAA CCGCAATTGG GCCCTTATCA ATCAGATTGA
151  TCACAGAAG AAAAAGACAT TCGGCATTCC AATGAAATCA CGCCTTTGGA
201  ATATAAAGGA TTAATTGGG CTGGCGTGGT GTTTGTGGCC TTATCGCCCC
251  TATTGGCTTG GAGCATCGTC CCGCCGACGC GTATTTCGCG TCATCCCTGAA
301  ACAGGATTGG TTTCGGGTTC GCGCTTTTAA AAATCGGATT TGTCTTTTAA
351  TTTCTGTGTT TTGCACATGC CGGCGATTGC TTATGCGCGT GTAACCCGAA
401  GTTTGCGCGG CGACACGAA GTGCTTAATG CGmGCGCGA ATCAGTAGAT
451  ACTCGGgGC TTnTTTgSw CakcATCTTT TTTGCCgCAC AGTTTGTGCG
501  ATTTTTTAAT TCGACGAATA TTGGGCAATA TATTGCCGTT AAAGGGCGGA
551  CGTTCCTTAAA AGAAGTCGGC TTGGGCGCGA CGGTGTGTT TATCGGTTTT
601  ATTTTAAATTT TGCCTTTTAT CAATCTGATG ATAGGCTCCG CCTCGCGCGT
651  ATGGGCGGTA ACTGCGCCGA TTTTCGTCCC TATGCTGATG TTGGCCGGGT
701  ACGCGCCCGA AGTCATTCAA GCGGCTTACC GATCGGTTGA TTCCGTTACC
751  AATATTATTA CGCCGATGAT GAGTTATTTC GGGCTGATTA TGGCGACGGT
801  GrkCmmnTAC AAAAAAGATG CGGGCGTGGG TaCGcTGATT wCTATGATGT
851  TGCCGTATTC CGCTTCTTTC TTGATTGCGt GGAATTGCCTT ATTCGTCATT
901  TGGGTATTG TTTTGGGCCT GCCCGTCGGT CCCGGCGCGC CCACATCTTA
951  TCCCGCACT TAA

```

This corresponds to the amino acid sequence <SEQ ID 134; ORF12>:

```

1  ..AXXIHXV VGPeanWFFM VASTFVIALI GYFVTEKIE PQLGPYQSDL
51  SQEEKDIRHS NEITPLEYKG LIWAGVVFVA LSALLAWSIV PADSLRHE
101  TGLVSGSEFL KSIIVLEFFL FALPSIVFGR VRSLRAGEE VPMXAXESL
151  TLXLXIXIF FAFGVFAFFN WINNGQYIAV KGATFFKEVG LGGSVLFIGF
201  ILICAFINLM IGSASQWAV TAPIFVPLML LAGYAEPIVQ AAYRIGDSVT
251  NIITPMMSYF GLINATVXXY KADAGVGTLI XMMLPYSAFF LIAWIALPCI
301  WVFVGLGLPVG PGATFYFAP *

```

Further sequence analysis revealed the complete DNA sequence <SEQ ID 135> to be:

```

1  ATGAGTCAAA CCGATACGCA ACGGGACCGA CGATTTTAC GCACAGTCGA
51  ATGGCTGGGC AATATGTTGC CGCATCCGTT TACGCTTTT ATTACTTTCA
101  TTGTGTTATT GCTGATTGCC TCTGCGTCG GTGCGTAITT CGGACTATCC
151  GTCCCCGATC CGCGCCCTGT TGTGCGGAAA GGACGTGCCG ATGACGGTTT
201  GATTACATT GTCAGCGTGC TCAATGCCGA CGGTTTTATC AAAATCCTGA
251  CGCATACGTT TAAAAATTC ACCGGTTTGG CGCGTTGGG AACGGTGTG
301  GTTTCCTTAT TGGCGGTGGG GATTGCGGAA AAATCGGGCT TGATTTCCGC
351  ATTAATGCGC TTATTGCTCA CAAATCGGCC ACGCAACTC ACTACTTTTA
401  TGGTGTGTTT TACAGGATT TTATCTAATA CCGCTCTGTA ATTTGGCTAT
451  GTGCTCGTAA TCCTTTGTCG CGCATCATC TTTCATCTCC TCGCGCGCA
501  TCGCTTGGC GGTCTGGCTG CGGCTTTCGC CGGCTTTCGC GCGGTTTATT
551  CGGCCAATCT GTTCTTAGGC ACAATCGATC CGCTCTGGC AGGCATCACC
601  CAACAGCGCG CGCAATCAT CCATCCCGAC TACGTCGTA GCCCTGAAGC
651  CAACTGGTGT TTTATGGTAG CCACTACGTT TGTGATTGCT TTGATTGGTT
701  ATTTTGTATC TGA AAAAATC GTGCAACCGC AATTGGGCCC TTATCAATCA
751  GATTGTGAC AAGAAGAAAA AGACATTGCG CATTCCAATG AAATCAGGCC
801  TTGGAATAT AAAGGATTA TTTGGCTGG CGTGGTGTG GTTGCCTTAT

```

851	CCGCGCCTATT	GCGTGGGAG	CGTGCCTCGT	CGCGCGGAT	TTTGGCTGAT
901	CCCTCAACAAG	GAGTGGTTTC	ACGTCGCGCC	TTTATAAAT	CGATTGCTAT
951	TTTATTATTTC	TGTGTTTTC	CAGTCGCGGG	CATTGTTTAT	GGCCGGGTAAC
1001	CCCGAAGTTT	CGCGGGCGAA	CAGGAAGCTG	TATAATCGAT	GGCCGAATCG
1051	ATGAGTACTC	TGGGGGCTTA	TTTGGCTCAT	ATCTTTTTTG	CGCGACAGTT
1101	TGTGCGATTT	TTTAAATGGA	GGAATATATG	CAATATTAAT	CGCGTTAAAG
1151	GGGCGACGCT	CTATAARAGA	GTCCGGCTTG	CGCGCAGCGT	GTGTGTTTATC
1201	GGTTTATTATT	TAAATTTGTC	TTTATCAAT	CTGATGATAG	CTGCGCGCTT
1251	CGCGCAATGG	CGGGTAACCT	CGCGCATTTT	CGTCCCTATG	CTGATGCTATG
1301	CGCGCTAGCG	CGCGGAGCT	ATTGACGCC	CTATACCAAT	CGCGTGATCC
1351	GTTATGCAAT	GATATGCTAT	CTATACCAAT	CGTATGCGG	CGATTATTGC
1401	GACCGTGATT	AAATACAAAA	AAGATACGGG	CGGGGTGAT	CTGATTTCTA
1451	TGATGTTCCT	TATCTCCGCT	TCTCTTCTGA	TGTGCTGGAT	TGCCCTTATT
1501	TGCATTTGGG	TATTTGTTT	GGAGCGTCCC	GTCCGTCCCG	CGCGCGCCAC
1551	ATTCTATCCC	GACCACTTA			

This corresponds to the amino acid sequence <SEQ ID 136; ORF12-1>:

	1	MSQDPTQRGD	RLRLTVEWIG	NMLPHVPVTF	ILFIVLLTLL	SAVGAYFGLS
	5	VEFDRPVQAG	GRADDGLIYI	VLNSLDAGFI	KITITVKNFK	TGFAPLGTVL
20	101	VSLLGVGIAE	KSLGLSALMR	LLLTQSPFRKL	TTFVMVFTGI	LSNTASELGI
	151	VVLVPLSAIT	FHSLIGRIHA	GAALAAFGVS	GGYANLFLG	TIDPLLAGIT
	201	QQAQAIIHPD	VYVGPPEANF	FMVASTETVA	LIGYFVTEKI	VEPQLPGQVS
	251	DLSQEKKDIR	HSNEITPLEY	KLLWAGVTFV	VALSALLASV	IVFADGILRH
	301	PETGLVGSQP	FLKSIVVFTF	LFALFALGIVY	GRVTRLSRG	QEVNMAAES
25	351	MSLTGLYLIY	LEFPAQEVAF	TRNTNGVQY	AVKGAFLGE	QVGGSGST
	401	GPILLINLQ	GLKLSAQWQ	AVYATFVFM	LLMAGYAPBV	IQAYRIGDS
	451	VTNILTFMMS	VEGLIDATVI	KPKKDAQVGT	LISMPLPSA	<u>FLFI</u> IAWIALE
	501	CIVWFYGLGL	VEGSGATTFV	AY*		

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

30 ORF12 shows 96.3% identity over a 320aa overlap with an ORF (ORF12a) from strain A of *N. meningitidis*:

[illegible]


```

orf12.pep  KKDAGVSTLIXMMLPYSAFFLIAMIALFCINWVFLGLPVGPGATFFYPAPX
           |||||
orf12a     KKDAGVSTLISMLPYSAFFLIAMIALFCINWVFLGLPVGPGATFFYPAPX
           480      490      500      510      520

```

5 The complete length ORF12a nucleotide sequence <SEQ ID 137> is:

```

1  ATGAGTCAAA CCGATACGCA ACGGACGCGA CGATTTTTC GCACAGTCGA
51  ATGGCTGGGC AATATGTTGC CGCACCGGTT TACGCTTTT ATTATTTTCA
101 TTGCTGTTAT GCTGATTGCC TCTGCCGCGG GTGCGTATT CGGACTATTC
151 GTCCCGGATC CGCGCCCTGT TGGTGCGAAA GGAACGTCCG ATGACGCTTT
201 GATTACAGTT GTACAGCTGC TCGATGCTGA CGGTTTGAT AAAATCCTGA
251 CGCATACCGT TAAAAATTTC ACCGGTTTCG CGCCGTGGG AACGGTGTG
301 GTTCTTTTAT TGGCGTGGG GATTGCGGAA AAATCGGGCT TGATTCCCG
351 ATTAATGCGC TTATTGCTCA CAAAATCTCC ACGCAAATC ACTACTTTTA
401 TGGTGTGTTT TACAGSGATT TTATCTAATA CGCTTCTGA ATTTGGGCTAT
451 GTCGCTCTAA TCCCTTTGTC CGCATCATC TTTCATCCC TCGCGCGCCA
501 TCCGCTTGCC GTCTGGCTG CGGCTTTGCG CGCGGTTTCG GCGGTTTAT
551 CGGCAATCT GTTCTTAGGC ACAATCGATC CGCTCTGGC AGGCATCAC
601 CAACAGGCGG CGCAAAATCAT CATCCCGGAC TACGCTGAG CGCTCAAGC
651 CAACTGGTTT TTATCGTAG CCACTACGTT TGTGATGCT TTGATGGTT
701 ATTTCTTAC TGA AAAAATC GTGAAACCG AATTGGCTCT TTATCAATCA
751 GATTGTGCAC AAGAAGAAAA AGACATTCGA CATTCCAAT AAATCACGCC
801 TTTGGAATAT AAAGATTAA TTTGGGCTGG CGTGGTGTG GTTGCCTTAT
851 CCGCCTTATT GCCTTGGAGC ATCGTCCCTG CGACGGTAT TTTGCCTCAT
901 CCGTGAACAG GATTGTTTTC CGGTCGCGG CTTTAAAT CAATTGTTGT
951 TTTTATTTTC TTGTGTTTTC CACTGCGCGG CATGTTTAT GGCCGGTAA
1001 CCGCAAGTTT CGCGCGCGAA CAGGAAGTCG TAAATCGAT GCGCGAATCG
1051 ATGAGTACTC TGGGCGCTTA TTTGGTCATC ATCTTTTTCG CCGCACAGT
1101 TGTGCGATT TTTAATTGGA CGAATATTGG CGAATATAT GCGGTTAAAG
1151 GGGCGACGTT CTTAAAGAA GTGCGCTTGG GCGCGACGCT GTTGTGTTATC
1201 GGTTTTATTT TAATTTTGTG TTTTATCAAT CTGATGATAG GCTCCGCTTC
1251 CGCGCAATGG GCGGTAACTG CGCGGATTTT CGTCCCTATG CTGATGTTGG
1301 CCGGCTACGC CGCGGAAGTC ATTCAAGCCG CTACCGCAT CGGTATGCTG
1351 GTTACCAATA TTATTACCC GATGATAGT TATTTGGGCG TGATATGCG
1401 GACGCTGATC AATACAAA AAGATCGCGG CGTGGGTACG CTGATTCTTA
1451 TGTGTTGCTC GTATTCCGCT TCTCTCTGTA TTGCGTGGAT TGCCTATTTC
1501 TGCATTGGG TATTGTTT GGGCTGCC GTCCGTCCG GCGCGGCCAC
1551 ATTCTATCCC GCACCTTAA

```

This encodes a protein having amino acid sequence <SEQ ID 138>:

```

1  MSQTDTRDGR FLRTVWEVLG NMLPHPVTLF IIFIVLLLIA SAAGAYFGLS
51  VDPDRFVGAK GRADDGLIHV VSLLDADGLI KILTHTVKNF TGFAPLGTVL
101 VSLLVGVIAE KSLGISALMR LLLTKSPRKL TTFMVVFTGI LNSTASELVI
151 VVLIPLSAII FHLGLRHLPA GLAAAFAGVS GGYSANLFLG TIDPLLAGIT
201 QQAQIIHPD YVVGPEANWF FMVASTFVIA LIGYFVTEKI VEPQLPGYTS
251 DLSQEEKDIR HSNEITPLEY KGLIWAGVVF VALSALLAWS IVPADGILRH
301 PETGLVSGSP FLKSIIVFIF LFLALPGIVY GRVTRSLRGE QEVVNMAES
351 MSTLGLYLVY IFFAAQGVAF FNWTNIGQYI AVKGATFLKE VGLGGSVLEFI
401 GFILICAFIN LMIQSASQW AVTAPLFVPM LMLAGYAEV IQAARYIGDS
451 VTNIITPMMS YFGLIMATVI KYKRDVAGVT LISMLPYSA FFLIAMIALF
501 CIWVFLGLP VGPGATFFY AF*

```

ORF12a and ORF12-1 show 99.0% identity in 522 aa overlap:

```

           10      20      30      40      50      60
orf12a.pep  MSQTDTRDGRFLRTVWEVLGNMLPHPVTLFIIFIVLLLIASAAGAYFGLSVDPDRPVGAK
           |||||
55  orf12-1   MSQTDTRDGRFLRTVWEVLGNMLPHPVTLFIIFIVLLLIASAVGAYFGLSVDPDRPVGAK
           10      20      30      40      50      60

           70      80      90      100     110     120
orf12a.pep  GRADDGLIHVSVLLDADGLIKILTHTVKNFTGFAPLGTVLVSLLVGVIAEKSLGISALMR
           |||||
60  orf12-1   GRADDGLIYVSVLLNADGFKILTHTVKNFTGFAPLGTVLVSLLVGVIAEKSLGISALMR
           70      80      90      100     110     120

```

		130	140	150	160	170	180
	orf12a.pep	LLLTSPKRLTTFMVVFTGLSNTASELGYVVLPLSAIIFHSLGRHPLAGLAAAFAGVS					
5	orf12-1	LLLTSPKRLTTFMVVFTGLSNTASELGYVVLPLSAIIFHSLGRHPLAGLAAAFAGVS					
		130	140	150	160	170	180
	orf12a.pep	GGYSANLFLGTIDPLLAGITQQAQIIHPDVVVGPEANWFFMVASTFVIALIGYFVTEKI					
10	orf12-1	GGYSANLFLGTIDPLLAGITQQAQIIHPDVVVGPEANWFFMVASTFVIALIGYFVTEKI					
		190	200	210	220	230	240
	orf12a.pep	VEPQLGPYQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFVLSALLAWSIVPADGILRH					
15	orf12-1	VEPQLGPYQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFVLSALLAWSIVPADGILRH					
		250	260	270	280	290	300
	orf12a.pep	PETGLVSGSPFLKSIVVFIFFLLFALPGIVYGRVTRSLRGEQEVVNMAESMSTLGLYLVI					
20	orf12-1	PETGLVSGSPFLKSIVVFIFFLLFALPGIVYGRVTRSLRGEQEVVNMAESMSTLGLYLVI					
		310	320	330	340	350	360
	orf12a.pep	IFFAAQFVAFNWTNIGQYIAVGKATFLKEVGLGGSVLFIFGILICAFINLMIGSASAGW					
25	orf12-1	IFFAAQFVAFNWTNIGQYIAVGKATFLKEVGLGGSVLFIFGILICAFINLMIGSASAGW					
		370	380	390	400	410	420
	orf12a.pep	AVTAPIFVPMMLAGYAPEVIAQAYRIGDSVTNIITPMMSYFGLIMATVIKYKKDAGVGT					
30	orf12-1	AVTAPIFVPMMLAGYAPEVIAQAYRIGDSVTNIITPMMSYFGLIMATVIKYKKDAGVGT					
		430	440	450	460	470	480
	orf12a.pep	LISMMPLPYSAFFLIAWIALFCIWWFVLGLFVGPAGTFYPAPX					
35	orf12-1	LISMMPLPYSAFFLIAWIALFCIWWFVLGLFVGPAGTFYPAPX					
		490	500	510	520		
	orf12a.pep	TLXLKXXIFFAAQFVAFNWTNIGQYIAVGKATFLKEVGLGGSVLFIFGILICAFINLM					
40	orf12-1	TLXLKXXIFFAAQFVAFNWTNIGQYIAVGKATFLKEVGLGGSVLFIFGILICAFINLM					
		490	500	510	520		

Homology with a predicted ORF from *N.gonorrhoeae*

ORF12 shows 92.5% identity over a 320aa overlap with a predicted ORF (ORF12.ng) from *N.*

45	<i>gonorrhoeae</i> :						
	orf12.pep			AXXIHPXXVVGPEANWFFMVASTFVIALI		30	
	orf12.ng	AAAFAGVSGGYSANLFLGTIDPLLAGITQQAQIIHPDVVVGPEANWFFMAASTFVIALI				232	
50	orf12.pep	GYFVTEKIVEPQLGPYQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFVLSALLAWSIV				90	
	orf12.ng	GYFVTEKIVEPQLGPYQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFVLSALLAWSIV				292	
55	orf12.pep	PADGILRHPEPETGLVSGSPFLKSIVVFIFFLLFALPGIVYGRVTRSLRGEQEVVNMAESMS				150	
	orf12.ng	PADGILRHPEPETGLVSGSPFLKSIVVFIFFLLFALPGIVYGRVTRSLRGEREVVNMAESMS				352	
	orf12.pep	TLXLKXXIFFAAQFVAFNWTNIGQYIAVGKATFLKEVGLGGSVLFIFGILICAFINLM				210	
60	orf12.ng	TLGLYLVIIFFAAQFVAFNWTNIGQYIAVGKATFLKEVGLGGSVLFIFGILICAFINLM				412	
	orf12.pep	IGSASAGWAVTAPIFVPMMLAGYAPEVIAQAYRIGDSVTNIITPMMSYFGLIMATVIXXY				270	
65	orf12.ng	IGSASAGWAVTAPIFVPMMLAGNAPQVIQAYRIGDSVTNIITPMMSYFGLIMATVIXY				472	

```

orf12.pep  KKDAGVGTGLIXMMLPYSAFFLIWIALFPCINWVFLVGLVPGFGPATFFYPAP 320
           |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
orf12.ng   KKDAGVGTGLISMMPLPYSAFFLIWIALFPCINWVFLVGLVPGFGPTFFYPVP 522

```

The complete length ORF12ng nucleotide sequence <SEQ ID 139> is:

```

5      1  ATGAGTCAAA  CCGACGCGCG  TCGTAGCGGA  CGATTTTTC  GCACAGTCGA
51     51  ATGGCTGGCG  AATATGTTCG  CGCACCCGGT  TACGCTTTT  ATTATTTTCA
101    101  TTGTGTTATT  GCTGATTGcc  cctgCCGTCG  GTGCGTATT  CGGACTATCC
151    151  GTCCCCGATC  CGCGTCCTGT  TGGGCGGAAA  GGACGTGCG  ATGACGGTTT
201    201  GATTACAGTT  GTCAGCCTCG  TCGATGCCGA  CGGTTTGATC  AAAATCCTGA
251    251  CGCATACCGT  TAAAAATTTC  ACCGGTTTCG  CGCGTGTGG  AACGCTGTTG
301    301  GTTCTCTTAT  TGGCGGTGGG  GATTGCGGAA  AAATCGGGCT  TGATTTCCGC
351    351  ATTAATGCGC  TTATTGCTCA  CAAAATCCCC  ACGCAAACCT  ACTACTTTTA
401    401  TGGTGTGTTT  TACAGGGATT  TTATCCAATA  CGGCTTCTGA  ATTGGGCTAT
451    451  GTCGTCTTAA  TCCCTTTGTC  CGCCGTCATC  TTTTCATCGC  TCGCGCGCGA
15     501  TCCGCTTGCC  GTTTGGCTGC  CGGCTTTTCG  CGCGGTTTCG  GCGGTTTATT
551    551  CGGCCAATCT  GTTCTTAGCG  ACAATCGATC  CGCTCTGGC  AGGCATCAC
601    601  CAACAGGCGG  CGCAAAATCAT  CCATCCGAC  TACCTGGTAC  GCGCTGAGC
651    651  CAACTGSGTT  TTTATGCGAG  CCACTACCTT  TGTGATTGCT  TTGATTGGTT
701    701  ATTTGTTTAC  TGAATAAATC  GTCCAGCCGC  AATTGGGCC  TTATCAATCA
751    751  GATTGTCCAC  AAGAAGAAAA  AGACATTCCG  CATTCCAATG  AAATCGACGC
20     801  TTTGGAATAT  AAGAGTATA  TTTGGGACG  CGTGGTGT  GTTGCCTTAT
851    851  CGCCCTTATT  GGCTTGGAGC  ATCGCTCCGT  CGACGGTAT  TTTGCGTCAT
901    901  CCTGAACAG  GATTGGTTGC  CGGTTCCGCG  TTTTAAAT  CGATTGTTGT
951    951  TTTTATTTTC  TTGTTGTTTG  CGCTGCCGGG  CATGTTTAT  GGCCGAGTAA
25     1001  CCCGAAGTTT  GCGCGCGGAA  CGGGAAGTCG  TTAATGCGAT  GGCCGATTCG
1051   1051  ATGAGTACTT  TGGGACTTTA  TTTGGTCATC  ATCTTTTTC  CGCCACAGTT
1101   1101  GTGCGCATTT  TTTAATTTGA  CGAATATTGG  GCAATATATT  GCGGTAAAG
1151   1151  GGGCGGTGTT  CTTAATAAGAA  GTCGCGCTGG  CGCGCAGTGT  GTTGTATTATC
1201   1201  GGTTTTATTT  TAATTGTGTC  TTTTATCAAT  CTGATGATAG  GCTCCGCTTC
30     1251  CGCGCAATGG  CGGATAACTG  CGCGGATTTT  CGTCCCTATG  CTGATGTGTTG
1301   1301  CCGGCTACGC  GCCCGAAGTC  ATTCAGCCGC  CTTACCGCAT  CGGTGATTCC
1351   1351  GTTACCAATA  TTATTACCGC  GATGATGAGT  TATTTCGGGC  TGATTATGCG
1401   1401  GACGSGTATC  AATATCAAAA  AAGATCGGG  CGTAGCGAG  CGATTATCGA
1451   1451  TGATGTCGTC  CGATTCGCTT  TTCTCTTAA  TTGCGTGGAT  CGCTTATTC
35     1501  TGCATTTGGG  TATTTCGTTT  GGGTCTGCC  GTCGCTCCG  GCACCCAC
1551   1551  ATTCTATCCG  GTGCCCTTAA

```

This encodes a protein having amino acid sequence <SEQ ID 140>:

```

1  MSQTDARRSG  RFLRTVWELG  NMLPHVPTLF  IIFIVLLLIA  SAVGAYFGLS
40     51  VPDPRFVGAK  GRADDGLIHV  VSLLDADGLI  KILHTVKNF  TGFAPLGTVL
101    101  VSLLVGVIAG  KSLGISALMR  LLLTKSPRKL  TTFMVVFTGI  LNSTASELGY
151    151  VVLIPLSAVI  FHSLSGRHPLA  GLAAAFAGVS  GGYSANLFLG  TIDFLAGIT
201    201  QQAQIIFHPD  YVVGPEANWF  FMAASTFVIA  LIGYFVTEKI  VEPQLGFYQS
251    251  DLSQEEKDIR  HSNEITPLEY  KGLIWAGVVF  VALSALLAWS  IVPADGLILH
301    301  PETGLVAGSP  FLKSIVVFIF  LLFALPGIVY  GRITRSLRGE  REVVNAMES
45     351  MSTLGLIYLI  IFFAAQFVAF  FWNINIGQYI  AVKAGVFLKK  FRLGSLVLEI
401    401  GFILICAFIN  LMIGSASQW  AVTAPIFVPM  LMLAGNAPOV  IQAAYRIGDS
451    451  VTNIITFMMS  YFGLIMATVI  KYKKDAGVGT  LISMMLPYSA  FFLIAWIALF
501    501  CINWVFLGLP  VPGPTTFYP  VP*

```

ORF12ng shows 97.1% identity in 522 aa overlap with ORF12-1:

```

50     100     200     300     400     500     600
orf12-1.pep  NSQTDTRQDRFLRTVWELGNMLPHVPTLFIIFIVLLLIASVAGVAYFGLSVDPDRFVGAK
           |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
orf12.ng     NSQTDARRSGRFLRTVWELGNMLPHVPTLFIIFIVLLLIASVAGVAYFGLSVDPDRFVGAK
           |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
55     70      80      90     100     110     120
orf12-1.pep  GRADDGLIYVSLNADGFIKILHTHTVKNFTGFAPLGTVLVSLVGLVGAIEKSGLSALMR
           |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
orf12.ng     GRADDGLIHVVSLLDADGLIKILHTHTVKNFTGFAPLGTVLVSLVGLVGAIEKSGLSALMR
           |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
60     130     140     150     160     170     180
orf12-1.pep  LLLTKSPRKLTTFMVVFTGILNSTASELGYVVLIPLSAIIFHSLGRHPLAGLAAAFAGVS

```

	orfl2ng	:	130	140	150	160	170	180
5	orfl2-1.pep		190	200	210	220	230	240
	orfl2ng		190	200	210	220	230	240
10	orfl2-1.pep		250	260	270	280	290	300
	orfl2ng		250	260	270	280	290	300
15	orfl2-1.pep		310	320	330	340	350	360
	orfl2ng		310	320	330	340	350	360
20	orfl2-1.pep		370	380	390	400	410	420
	orfl2ng		370	380	390	400	410	420
25	orfl2-1.pep		430	440	450	460	470	480
	orfl2ng		430	440	450	460	470	480
30	orfl2-1.pep		490	500	510	520		
	orfl2ng		490	500	510	520		

40 In addition, ORF12ng shows significant homology with a hypothetical protein from *E.coli*:

	sp P46133 YDAH ECOLI HYPOTHETICAL 55.1 KD PROTEIN IN OGT-DBPA INTERGENIC REGION
	>gi 1787597 (AE000231) hypothetical protein in ogt 5'region [Escherichia coli]
	Length = 510
	Score = 329 bits (835), Expect = 2e-89
45	Identities = 178/507 (35%), Positives = 281/507 (55%), Gaps = 15/507 (2%)
	Query: 8 RSGRFLRTVEWLGNMPLPHPVITXXXXXXXXXXASAVGAYFGLSVDPDRPVGAKGRADDGL 67
	Sbjct: 13 QSGKLYGWVERIGNKVPHPFLFYILIVLMVTTALSAFVSANKP-----FDGTP 64
50	Query: 68 IHVSLDADGLIKILTHTVKNFTGFAPXXXXXXXXXXIAEKSLGISALMRLLTKSP 127
	Sbjct: 65 VVVKNLLSVEGLHWFLPNVTKNFSGFAPLGAILALVLGAGLAERVGLLPALVMKMAASHVN 124
55	Query: 128 RKLTTFMVVFTGILSNTASELGYYVLIPLSAVIFHSLGRHPLAGLAAAFAGVSGGYSANL 187
	Sbjct: 125 ARYASVMVLTAFSSHISDAALVIMPFGALIFLAVGRHPVAGLLAAIAGVCGGFTANL 184
60	Query: 188 FLGTIDPLLAGITQQAQIIHPDYVVGPEANWFFMAASTFVIALIGYFVTEKIVEPQLGP 247
	Sbjct: 185 LVTVDVLLSGISTEAAANFPQMHVSVINDWYFMASVVVLTIVGGLITDKITPEPLGQ 244
	Query: 248 YQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFALSALLAWSIVPADGLRHPEPTGLVA 307
	Sbjct: 245 WQNSDEKQLTLETESQRF-----GLRIAGVVSLLFIAIALMVPQNGILRDPINHTVM 298
65	Query: 308 GSPFLKSIIVVFIFLLFALPGIVYGRITRSLRGEREVVNMAESMSTLGLYLXXXXXXXXX 367
	SFF+K I V I L F + + YG TR++R + ++ M E M + ++

Sbjct: 299 PSPFIKGIVPLIILFFVVSLAYGIATRTIRQADLPHLMIEPMKEMAGFIVMVFLAQF 358
 Query: 368 XXXXNWTNIGQYIAVKGAVFLKEVGLGGSVLFIFGILLICAFINLMIGSASQAQWAVTAPIF 427
 NR+N+G++I+V L+ GL G F+G L+ +F+ + I S SA W++ APIF
 Sbjct: 359 VAMFNWSNMKGFI+VAGLTDILESSGLSGIPAFVGLALLSSFLCMFIASGSIAWSILAPIF 418
 Query: 428 VFMLMLAGYAPEVIOAAYRIGDSVTNITPMMSYFGLIMATVIKYYKKDAGVGTLLISMMLP 487
 VFM ML G+ P Q +RI DS + P+ + L + + +YK DA +GT S++LP
 Sbjct: 419 VFMFMLLGHPAFAQLFRIADSSVLELAPVSPFVLEFLGLQRYKPDAGLGTYYSLVLP 478
 Query: 488 YSAFFLI+AWIALFCI+VWFVGLPVGPG 514
 Y FL+ W+ + W +++GLP+GPG
 Sbjct: 479 YPLIFLVVWLLMLLAW-YLVGLPIGPG 504

Based on this analysis, including the presence of several putative transmembrane domains and the predicted actinin-type actin-binding domain signature (shown in bold) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 17

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 141>:

```

1  ..ACAGCCGCGC CAGCAGGTTn CnCGGCTTTC GTTTTCGTAA CGGACAGTCn
51  GGTGGAGGTG TTCCGGACACA TCCAGACCGC AGTGGAAACA GSTTTTTTTC
101  ATGCGATTTC GGTTCCTCT GTTTTGGTG CGGCGGCACA AGACTCGGCA
151  ATgCCTTCGC GCACTCGGTC TATACCGGTA TTTTCAGCAA CGGAATGCG
201  GACGCGGyCA ATTTTCCCG CAGCGTCGCG CCATATGCC GTGTTTgTt
251  CTTCAGACGG CAGCAGGTG GTTTTGTGT ACACCTTgAT GCACGGAAaTA
301  TCCGCGGCAT GGAATTTCTTG CAGTACGTTT TCCACGTCTT CAATCTGCTG
351  TCCGCTGTTC GGAGCGCGCG CATCGACGAC GTGCAGCAGC ACATCgGcTT
401  gCGCGGTTTC TTCCAGCGTG GCgGAAAAG CGGAANTCAG TTTgTGCgCG
451  agATyGCTnA CGAATCCGAC GGTATCGGTC AGGAATATGC TGCATTGCGG
501  ACT..

```

This corresponds to the amino acid sequence <SEQ ID 142; ORF14>:

```

1  ..TAGAAGXXVF VFVTDQVEV FGNIQTA+VET GFFHGISVSS VFGAA+QD+SA
51  MASRSASIPV FSATEMR+AA IFFAASR+HP VECSSDGSRS VLLY+LMHGI
101  SPANISCTF STSSICPLF GAAASTTCS TSACAVSSV AEKAE+SLCG
151  RKLTNPTVSV RIMLHSG..

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF14 shows 94.0% identity over a 167aa overlap with an ORF (ORF14a) from strain A of *N.meningitidis*:

	orf14.pep		10	20	30
			TAGAAGXXVFVFVTDQVEVFNGNIQTAVET		
			: : : : : : :		
45	orf14a	GRQLGFLRVGGALFVITAAQARVNNALCDCLTTGAAGFAVFVFTDGMQVFGNVQPAVET	150	160	170
			180	190	200
	orf14.pep		40	50	60
			GFFHGISVSSVFGAAQDSAMASRSASIPVFSATEMR+AAIFFAASR+HPVFCSSDGSRS		
			: : : : : : : : : : : :		
50	orf14a	GFFHGISVSSVFGAAQDSAMASRSASIPVFSATEMR+AAIFFAASR+HPVFCSSDGSRS	210	220	230
			240	250	260

100 110 120 130 140 150
 orf14.pep VLLYTLMHGISPAWISCSCTFTSTSSICCLPLFGAAASTTCSSTACAVSSSSVAEKAEISLGC
 |||||
 orf14a VLLYTLMHGISPAWISCSCTFTSTSSICCLPLFGAAASTTCSSTACAVSSSSVAEKAEISLGC
 270 280 290 300 310 320
 160
 orf14.pep RXLTNPTVSVRIMLHSG
 |||||
 orf14a RSLTNPTVSVRIMLHSGLMYSRRVSSVAKSWFAYMFDVLVSRNLRLDLPTLVX
 330 340 350 360 370 380

The complete length ORF14a nucleotide sequence <SEQ ID 143> is:

	1	ATGAGGAGGATT	TGCAGGAAGAT	CGSGTTCGAT	GTCCGCGCGCG	TAAAGGTTAGG
15	51	TGCGCAGAGCGC	GAACTATCATC	GTTCGCATCAT	TCCCGCCGCGC	GGCAACAGCGG
	101	TGCGCAGAGCGA	TGAGGTCGTTT	CGGTTCTTTT	TGATGTGGCGG	CTTCGATTTT
	151	TTGCGCATGCA	TAGGGTGTGCG	GGGTGTGACCG	TATGTCCTGG	ATTTTCAACA
	201	GAATGTTCGGA	AAGTCGGATT	TTCGGTGTGG	CCGACGACGAC	CGCGCAGCGG
	251	TGCTCGGCTGT	AAATGAGGTC	GATCGCGTAG	ACCGCGCTGG	TACGCAGAAAG
20	301	CTGCTGCTTGC	ATCAGCCGGA	CGCAGCGCGG	CGAGGTGATG	CGCCGACAGA
	351	TTTAAACACG	CTCGCGCGTG	CGCGCTGTGG	TTTCCACAATA	CGCGCAGCTG
	401	ACTTCGCGCA	GGTGCACACG	CGGGATTGTG	TGAGAGTTAT	CTTGGGCGCT
	451	CAGCTCGGTT	TGTCGCTGAT	CGCGCGTGGG	TGTGTTGTAA	TAACTGCCCA
	501	AGCCCGCGTC	AATAATGCTT	TGTGCGACTG	CTCGACAAGC	GGCGCAGCAG
25	551	GTTTCGCGGT	GTTCGTTTTC	TTTACGGAAG	GTGATAGTCA	TGTTTCGGG
	601	AACGTCACAG	CGCGAGTGA	AACAGTTTGT	TTTCAATGGC	TTCGGTTTC
	651	GTCTGTGTTT	TGTCGGCGCG	GCATATATCT	GGCAATGGTA	TGCGCAGCTG
	701	CGTCTATACG	GGTATATAT	GCACAGGAAA	TGGCGACGCG	GGCAATTTT
	751	CGCGCAGATG	CGCGGCATAT	CGCGGTGTGT	TGTTTCTTCG	CGGCAAGTGG
30	801	GTGGGTTTGT	TGTACACACT	TGATGACRAG	AATATACGCG	GTATGATGCG
	851	CTCGACGTAC	CGAGTGTGAG	GTGTGTCCGT	GTGTGGAGCG	GTGTGGAGCG
	901	CGCGCATGCA	CGAGTGTGAG	CAGACATATG	CGTTTGGCGG	TTTCTCCAG
	951	CGTGGCGAAA	AGGCGCGAAA	TACGTTTGTG	CGGACATATG	CTGACGAATT
	1001	CGACGCTATG	CGTCAGGATA	TGCTGCTATT	CGGCACTGAT	GCACGACCGC
	1051	CGCGCGCTGC	TGTCAGGTTG	GGCGAAGACG	TGGTCTTTTC	CTATATGCCC
35	1101	CGACTTGGTC	AGGCGGTTGA	ACAGACTTGA	TGTTCCGACA	TTGGTATATG

This encodes a protein having amino acid sequence <SEQ ID 144>:

40

1	MEDLQSIGFD	VAAYKVGQR	EHRRLHHPFG	GNGEADDVLV	AFFLVGGDFD
51	LEFVGCGVA	YLDFPQGVG	AAARVARVPD	AAARVARVLE	DADDVAQTCK
101	LLFDFQDPAGG	ADGDAAE+NR	LAARCVGFKH	VLDFPGQGVQ	ADLVEDFLGR
151	QLGFLRVAGG	LGVAITAA+NR	NNALCDLCTF	GAGLFAVFVF	VTDGQMVGFG
201	NQVPVAFETG	FHGISVSSVF	GAAQYASAMA	SRSSASIPVS	ATEMRTAAFT
251	PAASRHMHPF	CSDSGDSRVL	LTLMHGISF	AMWSCSTFST	SSTCCPLPGLA
301	AASTTSCSTS	ACAVSSSSVS	KAEISLCLGRS	LTNPTVSVRI	MLHSGLMYSR
351	RAVVSSVAKS	WSFAYMEDIV	SRILNRDLDT	LV*	

It should be noted that this sequence includes a stop codon at position 118.

Homology with a predicted ORF from *N.gonorrhoeae*

ORF14 shows 89.8% identity over a 167aa overlap with a predicted ORF (ORF14.ng) from *N.*

gonorrhoeae:

50	orf14.pep	TAGAAGXVVFVEVTDSQVEVFGNIQTAVET	30
	orf14.ng	GRQGFFFRVGGSFVITAQAIGIDALCDCLTADAGFAVPAFVADGQMVGFGNVQPAVET	208
	orf14.pep	GFFHGISVSSVFGAAAQDSAMASRSASIPVFSATEMRTAAIPPAASRHMFPVCSSDGSRS	90
55	orf14.ng	GFFHGISVSSVFGAAQQSAMASRSASIPVFSATEMRTAAIPPAASRHMFPVCSSDGSRS	268
	orf14.pep	VLLYTLMHGISPAWISSCTFTSSICCPFLGAAASTTCSSTACAVSSSVKAEKAEISLGG	150
60	orf14.ng	VLLYTLMHGISAWIWISSCTFTSSICCPFLRAAASTTCSSTACTVSSKVAEKAEISLGG	328

```

orf14.pep      RKLNTPTVSVRIMLHSG      167
               | | | | | | | | | | |
orf14.ng       RSLTNTPTVSVRIMLHAGLMYSRRRAVSVRVAKSWSFYAMPDLVSRNLRLDLPTLV 382

```

The complete length ORF14ng nucleotide sequence <SEQ ID 145> is predicted to encode a protein

5 having amino acid sequence <SEQ ID 146>:

```

1  MEDLQEIIGFD VAAVKVGRQR EHHRLHMTQS GNGKADDVLF AFFLVGGDFD
51  LRVIGCGGVA CLPDFQCNVVG EADFAVVDDE AAVRAVIEV DADDAVCAQK
101  LLFDQPDAGG AGNAAEHQHC FVRAINGEHR VGLDFGQVVQ ADLVEFDFLGR
151  QGFQFRVGGG SFVITAQAGI DDALCDCLTA DAAGFAVFAF VADGQMDFVFG
201  NVQPAVETGF FHGISVSSVF GAAAQYSAMA SRASIPVFS ATEMRTAAIF
251  PAASRHMPEVF CSSDGSRSVL LYTLMHGISW AWISCSFST SSICCPPLFA
301  AASTTCSSTS ACTVSSKVAE KAEISLCGRS LTNPTVSVRI MLHAGLMYSR
351  RAVSVRVAKS WSFYAMPDLV SRLNRLDLPT LV*

```

Based on the putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 18

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 147>:

```

1  ..GCCCAITACT CCGACCGCAC TTGGAAGCCG CGTTTGGNCG CGCGCGGTCT
51  GCGGTATCTG CTTTATGGCA CGCTATTGCG GSTATTGTGT ATGATTTTGA
101  TCCCGCACTC GGGCAGCTTC GTTTCGCGCT AGCGGTGCGT CGCGCGCTTG
151  TCGTTCGCGC CCGTGTATGAT TCGCGTGTGA GAGTGTGCGT CAAATATGGC
201  GATCGAGCCG TTTAAGATGA TCGTGGCGCA CATGTGTAAC GAGGAGCAGA
251  AAA.NTACGC CTACGGGATT CAAAGTTTCT TAGCAATATC GGGCGCGGTC
301  GTGGCGGCGA TTCTGCGGTT TGTGTTGGG TATATCGGTT TGCGCAACAC
351  CGCCGANAAG GCGGTTGTGC CGCAGACCGT GGTGTCGGCG TTTTATGTGG
401  GTGCGCGGTT GCTGGTGATT ACCAGCGCGT TCACGATTTT CAAAGTGTGG
451  GAATACGANC CGGAACCTTA CGCCCGTTAC CACGGCATCT ATGTCGCGCG
501  GAATCAGGAA AAAGCCAAC TGATGCGACT CTTAAAA.CC CGCG..

```

30 This corresponds to the amino acid sequence <SEQ ID 148; ORF16>:

```

1  ..GHYSDRTWKP RLXGRRLPYL LYGTLIAVIV MILPNPNSGSF GFGYASIALAL
51  SFGLMIALL DVSNMAMQP FKMMVGDMMN EEQKXYAYGI QSFANTGAV
101  VAAILPFVFA YIGLANTAKK GVVEQTUVVA FYVGAALLVI TSAFTIFKVK
151  EYXFETYARY HGIDVAANQE KANWIALLXK A..

```

35 Further work revealed the complete nucleotide sequence <SEQ ID 149>:

```

1  ATGTCGGAAT ATAGCCTCA AACAGCAAAA CAGGTTTGGC CGCGCTGGCG
51  AAAAAGCAGC ATTTGGATGC TCAGTTTCGG CTTCTCGCG GTTCAGACGG
101  CCTTTACCTT GCAAAGCTCG CAATAGAGCC GCATTTTCCA AACGTCAGCG
151  GCAGACCCCG ACAATTTGGG CTGGTTTTTC ATCCTGCGCG CGCTGGCGGG
201  GATGCTGGTG CAGCCGATTG TCGGCCATTA CTCGACGCG ACTTGGAAAGC
251  CGCGTTTGGG CGGCCGCGGT CTGCCGTATC TGCTTTATGG CACGCTGATT
301  CGGTTTATFG TGAATGATTT GATGCCGAAC TCGGCGAGCT TCGGTTTCGG
351  CTATGCGTCG CTGGCGGGCT TGTGTTTCGG CGCGCTGATG ATTGCGCTGT
401  TAGACCTGTC GTCAAAATATG GCGATGCGAG CTCTTAAAGAT GATGCTCGCG
451  GACATGGTCA ACGAGGAGCA GAAAGGCTAC GCCTACGGGA TTCAAAGTTT
501  CTTAGCAAAAT ACGGGCGCGG TCGTGGCGGC GATTCGCGCG TTTGTGTTTG
551  CGTATATCGG TTTGGCGAAC ACCGCCGAGA AAGCGCTGTG CGCGCAGACC
601  GTGGTCGTGG CTTTATTATG GGGTGGCGCG TTGCTGGTGA TTACGAGCGC
651  GTTCACGATT TTCAAAGTGA AGGAATACGA TCCGGAAACC TACGCCCGGT
701  ACAAGCGCAT CGATGCGCG CAGAGTCAGG AARAGGCAAA TCGATCGGAA
751  CTCTTGAAAAC CGCGCCTTAA GCGGTTTATC CGCTTACTCT TGGTCAATAT
801  CTTCCTCTGG TTGCGCTTCC AATATATGTG GACTTACTCG CGAGCGCGCA
851  TTGCGGAAAA CGCTGCGCAC ACCACCGATG CGTCTTCGAT GGTATATCAG
901  GAGCGCGGTA ACTGATACGG CGTTTGGCGG GCGGTGCGAT CGGTTGCGCG

```

5

951	GTGATTTGTT	TGCTTTTGAT	TGGCGAAAGT	CGCCGATAAA	TACCATAGGT
1001	CGGGTTATT	CGGCTTTTG	TGCTTGGGCG	TGCTCGGCTT	TTTCTCCGTT
1051	TTCTTCATCG	GCACCAATA	TGCGCTGTGT	TGTGTTTATA	CCTTAATCGG
1101	CATCGCTTGG	CGGCGCMTA	TCACTTATCC	GTGACGATT	GTGACCAACG
1151	CCCTTGTGGG	CAGACATATG	GGCACTTACT	TGGGCTTTGT	TAACGCTGCT
1201	ATCTGATCG	CTCAAAATCG	TGCTGCTTCTG	TGCTGCTGCT	TGCTGCTGCT
1251	TATGCTGGCG	CTTCTTCAGG	CTTTGTGGTG	TGGCTGCTCC	TGGCTGCTCC
1301	TGCTCTGGCT	CGCGTTTCC	TGCTTCCGTA	TTAAAGAAAC	ACACGGCGGG
1351	TTTGG				

This corresponds to the amino acid sequence <SEQ ID 150; ORF16-1>:

1	MSEYTPQTAK	UGFLPALAKST	ITWMLSGFGLG	VQTAFTPLQSS	QMSRIFOTLG
5	ADPLKAGWFF	LGLPAGMLV	QVGVHYSYDR	TPQLKGRGR	LEPLYLTYLG
151	AVLWMLPMSN	SGSGFGFYAS	LAALSPGALM	LAITLDVSNM	AMQPFKMWG
151	DMNTEWQEG	AGYIGQSLAN	GRGVAVALIT	FVFVYJALIT	PAKGVHPOT
201	VVVAFVYGAA	LVLTSAFTI	FKVKEVDEPT	YARGYHIDVA	ANQKNWIE
251	LKLTAKPAFA	TVLTFQPCF	PAQYMNWYS	AGAEATENWV	TDKASSVWG
301	EAGNWYGLVA	AVGSAVAVIC	SFVLKAPVKN	YHNAQVFGCL	ALGALGFVSV
351	FIGNQVALV	LSYTTIGLIT	AGIITPYITI	YKALSGKHM	GTYIGLNGS
401	ICMPOIVASI	LSFVLFPMLG	GLOATMFLVG	GVWLLGAFS	VFLKETHGG
451	V*				

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF16 shows 96.7% identity over a 181aa overlap with an ORF (ORF16a) from strain A of *N. meningitidis*.

orf16.pep 10 20 30
 GHYSDRTWKPKRLGRRRLPEYLLYGTLLIAVIV
 orf16a 1FQTLGADPHSLGWFFLLPLLAGMLVQPIVGHYSDRTWKPKRLGRRRLPEYLLYGTLLIAVIV
 50 60 70 80 90 100
 orf16.pep 40 50 60 70 80 90
 MILMPNMSGSGFGFYASIALAALSGFALMIALLDVSNMAMQPFKMMVGMVMNEEQKXYAYGI
 orf16a MILMPNMSGSGFGFYASIALAALSGFALMIALLDVSNMAMQPFKMMVGMVMNEEQKGYAYGI
 110 120 130 140 150 160
 orf16.pep 100 110 120 130 140 150
 QSFLANTGAVVAAILPFVFAYIGLANTAKKGVVQPVVVVAFYVGAALLVITSAFTIEFKVK
 orf16a QSFLANTGAVVAAILPFVFAYIGLANTAKKGVVQPVVVVAFYVGAALLVITSAFTIEFKVK
 170 180 190 200 210 220
 orf16.pep 160 170 180
 EYXFETYARYHGIDVAANQEKANWIALKXA
 orf16a EYNFETYARYHGIDVAANQEKANWIELLKTAPEKAFWTVTLVQFFECWFAYQYMWTSAGAI
 230 240 250 260 270 280
 orf16a FGNVWHHTDASSVGVQEAGNWGVGLAAVQSVAAVISCVELAKVPKNKYHKAGYFAGCLALGA
 290 300 310 320 330 340

The complete length ORF16a nucleotide sequence <SEQ ID 151> is:

451 GACATGGGTCA ACGAGGAGCA GAAAGGCTAC GCCTACGGGA TTCAAAGTTT
 501 CTTAGCGAAT ACGGCGCGCG TCGTGGCGCG GATTCTGCGG TTTGTGTTTG
 551 CGTATCTCGG TTTGGCGAAC ACGCGCGAGA AAGGCGTTGT GCGCGAGACC
 601 GTGGTCTGCG CGTTTATATG GGGTGGCGCG TTGCTGTGTA TTACGAGCCG
 5 GTTCACGATT TTCAAAGTGA AGGAATACAA TCCGGAACCC TACGCCCGTT
 701 ACCACGGCAT CGATGTGCGC GCGAATCAGG AAAAAGCCAA CTGGATCGAA
 751 CTCTTGA AAA CCGCGCCTAA GCGGTTTGG ACGGTTACTT TGGTGC AATT
 801 CTCTGCTGG TTGCGCTTCC AATATATGTG GACTTACTCG GCAGGCGCGA
 101 TTGCGGAAAA CGTCTGGCAC ACCACCGATG CGTCTCCGT AGGTATCTAG
 10 901 GAGGCGGGA ACTGGTACCG CGTTTGGCG GCGGTGCGAT CGGTTGCGCG
 951 GGTGATTTGT TCGTTTGTAT TGGCGAAAGT CCGGAATAAA TACCATAAGG
 1001 CGGTTTATT CGGTGTTTG GCTTTGGCGG CGCTCGGCTT TTCTCGGTT
 1051 TTCTTCATCG GCAACCAATA CGCGCTGGTG TTGCTTATA CTTAATCGG
 1101 CATCGCTTGG GCGGCGATTA TCACCTATCC GCTGAGGATT GTGACCAACG
 15 1151 CTTTGTGCGG CAGCATATG GCGACTTACT TGGGCGTGT TAACGGCTCTG
 1201 ATCTGTATCG CGCAATCGT CGCTTCGCTG TTGAGTTTGG TCGCTTTCC
 1251 TATGCTGGCG GCGTTGACG CCACTATGTT CTTGGTAGGG GCGTGGTCC
 1301 TGCTGCTGGG GCGGTTTTCC GTGTTCTGTA TTAAGAAAC ACACGCGCGG
 1351 GTTTGA

This encodes a protein having amino acid sequence <SEQ ID 152>:

1 MSEYTPQTAK QGLPALAKST IWMLSFGLG VQTAFTLQSS QMSRIFQTLG
 51 ADPHSLGWFF ILPPLAGMLV QPIVGHYSDR TWKPRLGRR LPYLLYGLTI
 101 AVIVMILMPN SSGSFGFYAS LAALSFGALM IALLDVSSNM AMQPFKMMVG
 151 DMVNEEQKGY AYGIOQFLAN TGAVVAAILP FVFAYIGLAN TAEKGUVVPQT
 201 VVVAFYVGAA LLVITSAPTI FKVKYENPET YARYHGDIVA ANQEKANWIE
 25 251 LLKTAAPKAFW TVTLVQFFCW FAFQYMWYTS AGAIAENVWH TTDASSVGYQ
 301 EAGNWWYGLA AVQSVAIVIC FVLAKVPNK YHKAGYGCGL ALGALGFFSV
 351 FFIGNQYALV LSYTLIGIAW AGIITYPLTI VTNALSGKHM GTYLGLENGS
 401 TCMPOIVASL LSFVLFMPLG GLQATMFLVG GVVILLGAFS VFLIKETHGG
 451 V*

ORF16a and ORF16-1 show 99.6% identity in 451 aa overlap:

		10	20	30	40	50	60
35	orf16a.pep	MSEYTPQTAKQGLPALAKSTIWMLSFGLGVQTAFTLQSSQMSRIFQTLGADPHSLGWFF					
	orf16-1	MSEYTPQTAKQGLPALAKSTIWMLSFGLGVQTAFTLQSSQMSRIFQTLGADPHNLGWFF					
		10	20	30	40	50	60
40	orf16a.pep	ILPPLAGMLVQPIVGHYSDRTWKPRLGRRLPYLLYGLTIAVIVMILMPNSGSFGFYAS					
	orf16-1	ILPPLAGMLVQPIVGHYSDRTWKPRLGRRLPYLLYGLTIAVIVMILMPNSGSFGFYAS					
		70	80	90	100	110	120
45	orf16a.pep	LAALSFGALMIALLDVSSNMAMQPFKMMVGDMDVNEEQKGYAYGIOQFLANTGAVVAAILP					
	orf16-1	LAALSFGALMIALLDVSSNMAMQPFKMMVGDMDVNEEQKGYAYGIOQFLANTGAVVAAILP					
		130	140	150	160	170	180
50	orf16a.pep	FVFAYIGLANTAEGKGVVPQTVVVAFYVGAAALLVITSAPTIKFVKYENPETYARYHGDIVA					
	orf16-1	FVFAYIGLANTAEGKGVVPQTVVVAFYVGAAALLVITSAPTIKFVKYENPETYARYHGDIVA					
		190	200	210	220	230	240
55	orf16a.pep	ANQEKANWIELLKTAPKAFWTVTLVQFFCWFAFQYMWYTSAGAIAENVWHTTDASSVGYQ					
	orf16-1	ANQEKANWIELLKTAPKAFWTVTLVQFFCWFAFQYMWYTSAGAIAENVWHTTDASSVGYQ					
		250	260	270	280	290	300
60	orf16a.pep	EAGNWWYGLAAVQSVAIVICSFVLAKVPNKYHKAGYGCGLALGALGFFSVFFIGNQYALV					
	orf16-1	EAGNWWYGLAAVQSVAIVICSFVLAKVPNKYHKAGYGCGLALGALGFFSVFFIGNQYALV					
		310	320	330	340	350	360
65	orf16a.pep	EAGNWWYGLAAVQSVAIVICSFVLAKVPNKYHKAGYGCGLALGALGFFSVFFIGNQYALV					
	orf16-1	EAGNWWYGLAAVQSVAIVICSFVLAKVPNKYHKAGYGCGLALGALGFFSVFFIGNQYALV					
		310	320	330	340	350	360

		370	380	390	400	410	420
5	orf16a.pep	LSYTLIGIAWAGIITYPLTIVTNALSGKHMGTYLGLFNGSICMPQIVASLLSFVLFPMIG					
	orf16-1	LSYTLIGIAWAGIITYPLTIVTNALSGKHMGTYLGLFNGSICMPQIVASLLSFVLFPMIG					
		370	380	390	400	410	420
	orf16a.pep	GLQATMFLVGGVLLGLGAFSVFLIKETHGGVX					
10	orf16-1	GLQATMFLVGGVLLGLGAFSVFLIKETHGGVX					
		430	440	450			

Homology with a predicted ORF from *N.gonorrhoeae*

- 15 ORF16 shows 93.9% identity over a 181aa overlap with a predicted ORF (ORF16.ng) from *N. gonorrhoeae*:

	orf16.pep	GHYSDRTWKPRXLGRRLPYLLYGTLIAIV	30
	orf16ng	HFSNARRRPAQFGLVFHPAAAGDAGSADSGYISDRTWKPRXLGRRLPYLLYGTLIAIV	131
20	orf16.pep	MILMNSGSGFGFYASLAALSFGALMIALLDVSSNMAMPFKMMVGMVMVNEEQKXYAYGI	90
	orf16ng	MILMNSGSGFGFYASLAALSFGALMIALLDVSSNMAMPFKMMVGMVMVNEEQKXYAYGI	191
25	orf16.pep	QSFLANTGAVVAAILPFVFAYIGLANTAXKGVPVQTVVVAFYGAALLVITSFTIFKVK	150
	orf16ng	QSFLANTDAVVAAILPFVFAYIGLANTAEKGVPVQTVVVAFYGAALLITSFTISKVK	251
	orf16.pep	EYXPETYARYHIGDIVAANQEKANWIALLKXA	181
30	orf16ng	EYDPETVARYHIGDIVAANQEKANWFELLKTAPKVFVTVVQVFCWFCAFMYMTYSAGAI	311

The complete length ORF16ng nucleotide sequence <SEQ ID 153> is:

	1	ATGATAGGGG	ATCGCCGCGC	CGGCAACCAT	TTCGGATTTT	CCAAAGCAAA
35	51	TACTTTTCAA	ATCAAAAAAA	AGGATTACT	TTATGTCGGA	ATATACGCCT
	101	CAACACAGCAA	AACAAGGTTT	GCCCGCGCCG	GCAAAAAGCA	CGATTGGGAT
	151	GTTGAGCTTC	GGCTATCTCG	GCGTTCAGAC	GGCCTTTACC	CTGCAAAAGT
	201	CGCAGATGAG	CCGCATTTTT	CAAAACGCTA	GCGCAGACCC	GCACAATTTG
	251	GGCTGGTTTT	TCATCTCGCC	GCCGCTGGCG	GGGATCGTGG	TTACGCCGAT
	301	AGTGGCTACT	ACTCAGACCG	CACCTTGGAA	GCGCGTTGG	GCGCGCCGCG
40	351	CCTGCCGAT	CTGCGTTTAC	GCACGCTGAT	TGCGGTCATC	GTGATGATT
	401	TGATGCCGAA	CTCGGGCAGC	TTCGGTTTCG	GCTATCGCTC	CTGCGCGGCC
	451	TTGTGCTTCG	GCGCGCTGAT	GATTCGCTGT	TTCGACGTGT	CTCGCAATAT
	501	GCGCATCGAG	CCGTTTACGA	TGATGCTGG	CGATTCGCTC	AACGACGAG
	551	AGAAAGCTA	CGCTACCGG	ATTCAAAGTT	TCTTAGCGAA	TACGACGCG
45	601	GTTGTGGCAG	CGATTCTGCC	GTTTGTGTTT	CGGTATATCG	GTTTGGCGAA
	651	CACCTGCCGAG	AAAGCGCTTG	TGCCACAAC	CGTGGTCGTA	GCATTCTATG
	701	TGGGTGCCGC	GTTACTGATT	ATTACCACTG	CGTTCACAA	CTCCAAAGTC
	751	AAAGAATACG	ACCCGGAAC	CTACGCCGCT	TACCACGGCA	TCGATGTCCG
	801	CGCGAATCAG	GAAGAAAGCA	ACTGTTTCGA	ACTCTTAA	ACCGCGCTTA
50	851	AAAGTGTTTG	GACGGTTACT	CGGTACAGT	TTTTCTGCTG	GTTCCGCTTC
	901	CGGTATATGT	GGACTTACTC	GCGAGGCGCG	ATTGACAGAA	ACGTCTGGCA
	951	CACACCGAT	GCGTCTTCG	TAGGCCATCA	GGAGCGGGC	AACCGGTACG
	1001	GCGTTTGGC	GCGGCTGTAG			

This encodes a protein having amino acid sequence <SEQ ID 154>:

55	1	MIGDRRAGNH	FGFSKANTFQ	IKKKDLLYVG	IYASNSKTRF	ARAGKKHDL
	51	VELRLSRSD	GLYPAKLAE	PHFNARRRP	AGFLVHFA	AGGDAGSAD
	101	SGYSDRTWK	FRLGRRRLPY	LLYGTLIAIV	VMLFMVSGS	FGFYASLA
	151	LSFGALMIAL	LDVSSNMAMP	PFKMMVGMV	NEEQKXYAYG	IQSPLANTDA
	201	VVAAILPFVF	AYIGLANTA	EKGVPVQTVV	AFYVGAALLI	ITSFTISKVK
60	251	KEYDPETVAR	YHIGDIVAAN	EKANWFELLK	TAPKVFVTVT	PVQFVCFWFA
	301	RYMWTYSAGA	TAENVVHTTD	ASSVGHQEAG	NRVGVLAIV*	

ORF16ng and ORF16-1 show 89.3% identity in 261 aa overlap:

		30	40	50	60	70	80
5	orf16-1.pep	MLSFGLFVQTAFTLQSSQMSRIFQTLGADPHNLGWFFILPPLAGMLVQPI-VGHYS	DRT				
	orf16ng	DVELLSRRSDGLYPAKLADEPHFSNARRRPAQGLVLF-HFAAAGGDAGSADSGYSDRT					
		50	60	70	80	90	100
10	orf16-1.pep	WKPRILGGRRLLPYLLYGTLLIAVIVMILMPNSGSGFGFYASLAALSFGALMIALLDVSSNMA					
	orf16ng	WKPRILGGRRLLPYLLYGTLLIAVIVMILMPNSGSGFGFYASLAALSFGALMIALLDVSSNMA					
		110	120	130	140	150	160
15	orf16-1.pep	MQPFKMMVGDVMVNEEQKSYAYGIQSFLANTDAVVAAILPFVFAYIGLANTAEGKVVPQTV					
	orf16ng	MQPFKMMVGDVMVNEEQKSYAYGIQSFLANTDAVVAAILPFVFAYIGLANTAEGKVVPQTV					
		170	180	190	200	210	220
20	orf16-1.pep	VVAFYVGAALLVITSFTIIFKVKYDEPTYARYHGIDVAANQEKANNIELLKTA					
	orf16ng	VVAFYVGAALLVITSFTIIFKVKYDEPTYARYHGIDVAANQEKANNIELLKTA					
		230	240	250	260	270	280
25	orf16-1.pep	VTLVQFFCWFPAFYQMMWTYSAGAIAENVWHTTDASSVGYCEAGNWWGVLA					
	orf16ng	VTLVQFFCWFPAFYQMMWTYSAGAIAENVWHTTDASSVGYCEAGNWWGVLA					
		290	300	310	320	330	340

Based on this analysis, including the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 19

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 155>:

```

1  ATGTTGTTCC  GTAAACGAC  CGCGCGCGTT  TTGGCGCAT  CCTTGATGCT
51  GAACGGCTGT  ACGTTGATGT  TGTGGGGAAT  GAACAACCCG  GTCAGCGAA
101  CAATCACCCG  NAACACCTTT  GNCAAGACC  AAATCCGNGN  CTTGCTGTG
151  GTTGCCGAAG  ACAATGCCCA  ATTGGAAGA  GGCACCTGG  TGATGATGG
201  CGGAAATATC  TGGTTCCTCG  TCAATCCCA  AGATTCCGCG  AA, NTGACGG
251  GNATTTTGAN  GGCAGGCGTG  GACAAACCT  TCCAATAGT  TNAGGATACC
301  CCGAGCATAT  C.TGCCACCA  AGCCCTGCG  GTCAAATCG  GATCGNCTGG
351  CAGCCGAAT...
```

This corresponds to the amino acid sequence <SEQ ID 156; ORF28>:

```

1  MLFRKTTAAV  LAHTLMLNGC  TMLWGMNPN  VSETITRKHV  XKDQIRXFGV
51  VAEDNAQLEK  GSLVMMGGKY  WEFVNPDESA  XXTGILXAGL  DKPFQIVXDT
101  PSYXCHQALP  VKLGSGSQN...
```

Further work revealed the complete nucleotide sequence <SEQ ID 157>:

```

1  ATGTTGTTCC  GTAAACGAC  CGCGCGCGTT  TTGGCGCAA  CCTTGATGCT
51  GAACGGCTGT  ACGTTGATGT  TGTGGGGAAT  GAACAACCCG  GTCAGCGAA
101  CAATCACCCG  CAACACCTTT  GACAAAGACC  AAATCCGCGC  CTTGCTGTG
151  GTTGCCGAAG  ACAATGCCCA  ATTGGAAGA  GGCAGCTGG  TGATGATGG
201  CGGAAATATC  TGGTTCCTCG  TCAATCCCA  AGATTCCGCG  AAGCTAGTG
251  GCATTTTGAA  GGCAGGCGTG  GACAAACCT  TCCAATAGT  TGAGGATACC
```

5
301 CCGAGCTATG CTGCGCACCA AGCCCTGCGG GTCAAACTCG AATGCGCTGG
351 CAGCGCAAT TTGCTTACCG AAGGCTTTTG CCTGCGCTAG GATACGACGA
401 AGCCTGCGCA CATCGCCAAG CTGAACACAG CTGGGTTTGA AGCGGTCAAA
451 CTGACACATC GGACCATTTA CACGCGCTGC GTATCGCGCA AAGGCAAAAT
501 CTACGCCACA CGCAAAAAC TGAACGCCGA TTACCATTTT GAGCAAAATG
551 TGCTGCGCGA TATTATTATC ACGGTTACTG AAGAACATAC CGACAAATCC
601 AAGCTGTTTG CAAATATCTT ATATACGCCC CCCTTTTGA TACTGGATGG
651 GGGCGGCGCG TGAATGGCTT TGCCTGCGGC GGCCTGGGT GCGGTCGTGG
701 ATGCCGCCCG CAAATGA

10 This corresponds to the amino acid sequence <SEQ ID 158; ORF28-1>:

1 MLFRKTTAAV LAATLMLNGC TMLWGMNPP VSETITRKHV DKDQIRAFGV
51 VAEDNAQLEK GSLVMMGGKY WVVNVPEDSA KLTGILKAGL DKPPQIVEDT
101 PSYARHQALP VKLESFPGSQN FSTEGLCRLY DTDKPADIAK LKQLGFEAVK
151 LDNRTIYTRC VSAKGGYYAT PQKLNADYHF EQSVADIIYY VTTEHTDKS
15 KLFANILYTP PFLILDAAGA VLALFAAALG AVVDAARK*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF28 shows 79.2% identity over a 120aa overlap with an ORF (ORF28a) from strain A of *N.*

meningitidis:

20
orf28.pep MLFRKTTAAVLAHTIMLNGCTIMLWGMNPFVSETITRKHVVKDQIRAFGVVAEDNAQLEK
orf28a MLFRKTTAAVLAATLMLNGCTVMWGMNPFSETTARKHVKDQIRAFGVVAEDNAQLEK
25
orf28.pep 70 80 90 100 110 120
GSLVMMGGKYWVVNVPEDSAXXTGILKAGLDRKPFQIVXDTPSYXCHQALPVKLSXGSON
orf28a 70 80 90 100 110 120
GSLVMMGGKYWVVNVPEDSAKLTLGILKAGLDRKPFQIVXQVVEFNFRFA-YQALPVKLESFASQN
30
orf28a FSTEGLCRLYDTRPADIAKLQLEFEFAVELDNRTIYTRCVSAKGGYYATPQKLNADYHF
120 130 140 150 160 170

The complete length ORF28a nucleotide sequence <SEQ ID 159> is:

35
1 ATGTTGTTC GTAAACGAC CGCGCGGCTT TTGGCGGCAA CTTTGATGTT
51 GAACGGCTGT ACGGTAATGA TGTGGGGTAT GAACAGCCCG TTGAGCGAAA
101 CGACCGCCCG CAACACAGTT GACAAGGACC AAATCCGCGC CTTCGGTGTG
151 GTTGCCGAAG ACAATGCCCA ATTGGAANAAG GGCAGCCTGG TGATGATGGG
201 CGGGAATAC TGGTTCGTG TCAATCCTGA AGATTGGCGC AAGCTGACGG
40 251 GCATTTTGAA GGCGGGGTTG GACAAGCAGT TTCAATAGGT TGAGCCCAAC
301 CCGCGCTTTG CTTACCAAGC CCGCGCGGTC AAACCTGAAAT CGCCGCGCAG
351 CCAGAAATTC AGTACCGAAG GCGTTTGCTT GCGTACGAT ACCGACAGAC
401 CTGCCGACAT CGCCAGCTG AAACAGCTTG AGTTTGAAAG GGTGGAACCT
45 451 GACATTCGCA CCRITTCAC GCGCTGCTC TCGCGCAAG GCAATACCTA
501 CGCACACCG CAAACACTGA ACCCGGATTA CCAATTTGAG CAAGTGTGC
551 CTGCGATAT TATTATCAC GTTACGAAA AACATACGCA CAATCCAG
601 TTGTTTGAAT ATATTGCATA TACGCCACAC ACGTTGATAC TGGATCGGCT
651 GGGCGCGGTG CTGGCGTTCG CTGTCGCGGC GTTGATTGCA GCCACGAATT
701 CCTCAGACAA ATGA

50 This encodes a protein having amino acid sequence <SEQ ID 160>:

55
1 MLFRKTTAAV LAATLMLNGC TVMWGMNPP FSETTARKHV DKDQIRAFGV
51 VAEDNAQLEK GSLVMMGGKY WVVNVPEDSA KLTGILKAGL DKQFMVEFN
101 PRFAYQALPV KLESFASQNF STEGLCLRDY TDRPADIAKL KQLEFEAVEL
151 DNRTIYTRC VSAKGGYYATP QKLNADYHFE QSVADIIYY VTTRKHTDKS
201 LFENIAYTPT TLILDAVGAV LALFAAALIA ATNSSDK*

ORF28a and ORF28-1 show 86.1% identity in 238 aa overlap:

		10	20	30	40	50	60	
	orf28a.pep	MLFRKTTAAVLAATLMLNGCTVMMGMNSPFSETTRARKHVDKQIRAFGVVAEDNAQLEK						
5	orf28-1	MLFRKTTAAVLAATLMLNGCTLMLGMNPNVSETITRKHVDKQIRAFGVVAEDNAQLEK						
		10	20	30	40	50	60	
	orf28a.pep	70	80	90	100	110	119	
	orf28-1	70	80	90	100	110	120	
10	orf28a.pep	GSLVMMGGKYWFVVPEDSAKLTGILKAGLDKQFQMFVFNPRFA-YQALPVKLESPASQN						
	orf28-1	GSLVMMGGKYWFVVPEDSAKLTGILKAGLDKPFQIVEDTFSYARHQALPVKLESPGSQN						
		70	80	90	100	110	120	
	orf28a.pep	120	130	140	150	160	170	179
	orf28-1	120	130	140	150	160	170	180
15	orf28a.pep	FSTEGCLCLRYDTRDPADIAKLQLEFEAVELDNRTIYTRCVSAKGKYYATPOKLNADYHF						
	orf28-1	FSTEGCLCLRYDTRDPADIAKLQLEFEAVELDNRTIYTRCVSAKGKYYATPOKLNADYHF						
		130	140	150	160	170	180	
	orf28a.pep	180	190	200	210	220	230	
	orf28-1	180	190	200	210	220	230	
20	orf28a.pep	EQSVPADIIYYTVTKKHTDKSLFENIAYTPTLLIDAVGAVLALPVAALIAATNSDDKX						
	orf28-1	EQSVPADIIYYTVTEHTDKSLFANILYTPFLILDAAGAVLALPAALGAVVDAARKX						
		190	200	210	220	230		

25 Homology with a predicted ORF from *N.gonorrhoeae*

ORF28 shows 84.2% identity over a 120aa overlap with a predicted ORF (ORF28.ng) from *N.*

gonorrhoeae:

	orf28.pep	MLFRKTTAAVLAHTLMLNGCTLMLWGMNPFVSETITRKHVVDKQIRAFGVVAEDNAQLEK	60
30	orf28ng	MLFRKTTAAVLAATLILNGCTMMLRGMNPFVSETITRKHVVDKQIRAFGVVAEDNAQLEK	60
	orf28.pep	GSLVMMGGKYWFVVPEDSAXXTGILKAGLDKPFQIVXDTPSYXCHQALPVKLGSKGSSQN	120
	orf28ng	GSLVMMGGKYWFVVPEDSAKLTGILKAGLDKPFQIVEDTFSYARHQALPVKFEAPGSQN	120

35 The complete length ORF28ng nucleotide sequence <SEQ ID 161> is

	1	ATGTGTTGCC	GTAACACGAC	CGCCGCGCTT	TTCGCGCGAA	CCTTGATACT
	51	GAACGGCTGT	ACGATGATGT	TGCGGGGGAT	GAACAACCCG	GTCAGCCAAA
	101	CAATCACCCG	CAACACCGTT	GACAAAGACC	AAATCCGCGC	CTCCGTTGTG
	151	GTTGCCGAAG	ACAATGCCCA	ATTGGAAGAG	GCGAGCCTGG	TGATGATGGG
40	201	CGGGAATATC	TGGTTCGCGC	TCAATCCCGA	AGATTCCGCG	AAGCTGACGG
	251	GCCTTTTGAA	GGCCGGGTTG	GACAAGCCCT	TCCAAATAGT	TGAGGATACC
	301	CCGAGCTATG	CCC GCCACCA	AGCCCTGCGC	GTCAAATTCG	AAGCGCCCGG
	351	CAGCCAGAAAT	TTCAAGTACG	GAGGTCTTTG	CCTCCGCTAT	GATACCGCGA
	401	GACCTGACGA	CATCGCCAA	CTGAACAGC	TTGAGTTTAA	ACCGGTCAAA
45	451	CTCGCAATCT	GGACCATTTA	CACCGCTGCG	GTATCCGCGA	AAGCAAAATG
	501	CTACGCCACG	CCGCAAAAC	TGAACGCCGA	TTATCATTTT	GAGCAAAATG
	551	TGCCCGCCGA	TATTTATTAT	ACGGTTACTG	AAAAACATAC	CGACAAATCC
	601	AAGCTGTTTG	GAAATATCTT	ATATACGCCC	CCCTTGTGTA	TATTGGATGC
50	651	GGCGGCCGCG	GTGCTGGTCT	TGCTATGGC	TCTGATTGCA	GCCGCGAATT
	701	CCTCAGACAA	ATGA			

This encodes a protein having amino acid sequence <SEQ ID 162>:

	1	MLFRKTTAAV	LAATLILNGC	TMLRGMNPF	VSQITIRKHV	DKDQIRAFGV
	51	VAEDNAQLEK	GSLVMMGGKY	WFAVNPEDSA	KLTLGLKAGL	DKPFQIVEDT
	101	PSYARHQALP	VKFEAPGSQN	FSTGGCLRLY	DTGRPDIAK	LKQLEFKAVK
	151	LDNRTIYTRC	VSARKGKYAT	POKLNADYHF	EQSVPADIIY	TVTEKHTDKS
55	201	KLFGNIIYTP	PLILDAALAA	VLVLFMALIA	AANSSDK*	

ORF28ng and ORF28-1 share 90.0% identity in 231 aa overlap:

	orf28-1.pep	10	20	30	40	50	60
60	orf28-1.pep	MLFRKTTAAVLAATLMLNGCTLMLWGMNPFVSETITRKHVVDKQIRAFGVVAEDNAQLEK					

	orf28ng	MLFRKTTAAVLAATLILNGCTMMLRGMNPNFSQTITRKHKVDKQIRAFGVVAEDNAQLEK	10	20	30	40	50	60
			70	80	90	100	110	120
5	orf28-1.pep	GSLVMGKGKWFVFNPEDSAKLTGILKAGLDKPFQIVEDTFSYARHQALFVKLESPGSQN						
	orf28ng	GSLVMGKGKWFVFNPEDSAKLTGILKAGLDKPFQIVEDTFSYARHQALFVKLESPGSQN	70	80	90	100	110	120
10	orf28-1.pep	FSTEGLCRLRYDTDKPADIAKLQGLFEAVKLDNRTIYTRCVSAKGKYYATPQKLNADYHF	130	140	150	160	170	180
	orf28ng	FSTGGLCLRYDTGRPDIAKLQGLFEAVKLDNRTIYTRCVSAKGKYYATPQKLNADYHF	130	140	150	160	170	180
15	orf28-1.pep	EQSVPADIIYTVTEHTDKSKLFANILYTPPELLDAAAGAVLALPAALGAVVDAARKX	190	200	210	220	230	239
20	orf28ng	EQSVPADIIYTVTEHTDKSKLFGNILYTPPELLDAAAVLVLPALIAAANSDDKX	190	200	210	220	230	

Based on this analysis, including the presence of a putative transmembrane domain in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF28-1 (24kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 6A shows the results of affinity purification of the GST-fusion protein, and Figure 6B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for ELISA, which gave a positive result. These experiments confirm that ORF28-1 is a surface-exposed protein, and that it may be a useful immunogen.

Example 20

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 163>:

	1	..GTCAGTCCTG TACTGCCTAT TACACACGAA CGGACAGGCT TTGAAGGTGT
35	51	TATCGGTAT GAAACCCATT TTTGAGGCA CGGACATGAA GTACACAGTC
	101	CGTTGCATCA TCATGATTCA AAAAGCACTT CTGATTTTCA CGGCGGTGTA
	151	GACGCGGTT TTACTGTTTA CCAACTTCAT CGAATATGTT CGGAATATCA
	201	TCCGAGGAT GAATATGACG GGCCGCAAGC AGCG.ATTAT CGCCCCCGC
	251	GAGGAGCAAG GGATATATAC AGCTATTATG TCAAAGGAAC TTCAACAAA
40	301	ACAAAGACTA GTATTGTCCC TCAAGCCCCA TTTTCAGACC GTTGCTAGA
	351	AGAAATGCC GGTGCCGCT CTGGT..

This corresponds to the amino acid sequence <SEQ ID 164; ORF29>:

	1	..VSPVLPITHE RTGFEGVIGY ETHFSHGHE VHSFPDHDS KTSDFSGGV
35	51	DGGFTVYQLH RTWSEIHPED ETDGPQAAKY PPGGARDIY SYVKGSTK
	101	TKTSIVPQAP FSDRWLEENA GAASG..

Further work revealed the complete nucleotide sequence <SEQ ID 165>:

	1	ATGAATTTCG CTATTCAAAA ATTCATGATG CTGTTTGCAG CAGCAATATC
35	51	GTTGCTGCAA ATCCCCATTA GTCATGCGAA CGGTTTGGAT GCCCGTTTCG
	101	GCGATGATAT GCAGGCCAAA CACTACGAAC CGGTTGGTAA ATACCATCTG
45	151	TTTGGTAATG CTCGCGCGAG TGTAAAGAG CGGTTTACG CCGTCCAGAC

The complete length ORF29a nucleotide sequence <SEQ ID 167> is:

1 ATGAATTNGC CTATTCAAAA ATTCAATGATG CTGTTTGCAG CAGCAATATC
51 GTNGCTGCAG ATCCCNATTA GTCATGCGAA CGGTTGGAT GCCGTTTGC
101 CGCATGATAT CGAGGCAAAA CACTACGAAC CGGTTGGTAA ATACCATCTG
151 TTTGGTAATG CTCGGCGCAG TGTAAAAAT CGGTTTACG CGGTCCAAAC
201 ATTTGATGCA ACTGCGGTCG GCCCATCTACT GCCTATTACA CAGCAACGAA
251 CAGGATTTGA AGGCATTATC GGTATTGAAA CCGATTTTTC AGCAATTCGA
301 CATGAAGTAC ACAGTCCCTT GATTAATCAT TATTCAAAA CACTCTCTGA
351 TTTCGTCGCA GGCCTAGAGC CTGGTTTATC CGTTTACCAA CTTCATCCGA
401 CAGGCTCGCA AATCCATCG GAGGATGGAT ATGACGGGCC GCAAGGCAGC
451 GATTATCCGC CCCC CGGAGG AGCAAGGAT ATATACANNT ANTATGCAA
501 AGGAACCTCA ACAAACAACA AGAGTAATAT TGTTCGCCGA GCCCATTTT
551 CAGACCGCTG GCTAAAGAAA AATGCCGGTG CCGCTCTGG TTTTTCAGC
601 CGTGCTGATG AAGCAGGAAA ACTGATATGG GAAAGCGACC CCAATAAAAA
15 651 TTGGTGGGCT AACCGTATGG ATGATATTGG CGGCATCGTC CAAGTGGCGG
701 TTAATCCTTT TTTAATGGGT TTTCAAGGAG TAGGATTTGG GGCAATTACA
751 GACAGTGCAG TAAGCCCGGT CACAGATACA GCCCGCAGC AGACTCTACA
801 AGGTATNAAT CATTTAGGAA ANTAAAGTCC CGAAGACAAA CTTCGGCTG
851 CAACCGCATT ACAAGACAGT GCTTTTGGCG TAAAGACGG TATCAATTCC
20 901 GCCAGACAA TGGCTGATGC CCAATCGAAT ATAACTGCAA CAGCCCAAC
951 TGCCCTTGCC GTAGCAGANG CCGCAACTAC GGTTTGGGGC GGTAAAAAAG
1001 TAGAACCTAA CCGACACAAA TGGGATTTGG TAAAAATAC NGGCTATAAN
1051 ACACCTGCTG TTCGCACCAT GCATCTTTTG GATGGGAAA TGGCCGGTGG
1101 GAATAGACCG CTTAATCTA TAACCTCCAA CAGCAACGCA GATGCTTCA
1151 CACACCCCTC TTACAGAGC CAACCTAATG GAGACAAAT TANNNNNGG
1201 CATGCTTATA ACAAGCATGT CATAAGACAA CAAAGATTTA CGGATTAAAA
1251 TATCAATTCA CACGACAGAT TTGCTCGGCA TATTGAAAT ATTGTTAGCC
1301 ATCCANCAAA TATGAAAGAG TTACTCTGCG GTAGAAGTGC GTATTGGGAT
1351 NATAAACAGC GGACNATAGT TATCCGAGAT AAAAATTCTG ACGATTGGAG
30 1401 TACAGCATT TACAGCAAT CAGGTAAAA ATATTATGAT GATTATATG

This encodes a protein having amino acid sequence <SEQ ID 168>:

1 MNXPIQKFM LFAAAISXKL IPISHANGLD ARLRDDMQAK HYPEGGKYHL
51 FGNARGSVKN RVYAVQTFDA TAVGPILPIT HERTGFEGII GYETHFSGHG
35 101 HEVHSPFDNH DSKSTSDFSG GVDGGFTVYQ LHRTGSEIHP EDGYDGPQGS
151 DYPPPGGARD IYXXYVKGTS TKTKSNIVPR AFPSDRWLKE NAGAASGFFS
201 RADEAGKLIW ESDPNKNWWA NRMDDIRGIV QGAVNPFIMG FQGVIGAIT
251 DSAVSPVTD TAAQTLQGXN HLGXLSPEAQ LAALALQDS AFAYKDGINS
301 ARQWADAHFN ITATAGTALA VAXAATTVWG GKVELNPTK WDWKNTGYX
351 TPAVRTMHTL DGEAGGNRP PKSITNSKA DASTQPSLQA QLIGEQIXX
40 401 HAYKNHVRQ QETFDLNNIS PADFAHLEN IVSPHNMKE LFRGRTAYWD
451 XKTGTIVIRD KNSDDGGTAF RPTSGKKYYD DL*

ORF29a and ORF29-1 show 90.1% identity in 385 aa overlap:

		10	20	30	40	50	60
45	orf29a.pep	MNXPIQKFM	LFAAAISXKL	IPISHANGLD	ARLRDDMQAKHYE	PGGKYHL	FGNARGSVKN
	orf29-1	MNLPIQKFM	LFAAAISL	QIPISHANGLD	ARLRDDMQAKHYE	PGGKYHL	FGNARGSVKN
		10	20	30	40	50	60
50	orf29a.pep	RVYAVQTFDA	TAVGPILPIT	HERTGFEGII	GYETHFSGHG	HEVHSPFDNH	DSKSTSDFSG
	orf29-1	RVYAVQTFDA	VSPLVPI	THERTGFEG	IVGYETHFS	GHGHEVHSP	FDNHDSKSTSDFSG
		70	80	90	100	110	120
55	orf29a.pep	GVDGGFTVYQ	LHRTGSEIHP	EDGYDGPQGS	DYPPPGGARD	IYXXYVKGTS	TKTKSNIVPR
	orf29-1	GVDGGFTVYQ	LHRTGSEIHP	EDGYDGPQGS	DYPPPGGARD	IYXXYVKGTS	TKTKSNIVPR
		130	140	150	160	170	180
60	orf29a.pep	AFPSDRWLK	ENAGAASGFFS	RADEAGKLIW	ESDPNKNW	ANRMDIRG	IVQAVNPFIMG
	orf29-1	AFPSDRWLK	ENAGAASGFFS	RADEAGKLIW	ESDPNKNW	ANRMDIRG	IVQAVNPFIMG
		190	200	210	220	230	240

		190	200	210	220	230	240
		250	260	270	280	290	300
5	orf29a.pep	FQGVGIGAITDSAVSPVTDAAQQTLOGXNHLGKLSPEAQLAAALQDSFAFAVKDGIN					
	orf29-1						
		FQGVGIGAITDSAVSPVTDAAQQTLOGINDLGKLSPEAQLAAALQDSFAFAVKDGIN					
		250	260	270	280	290	300
10	orf29a.pep	ARQWADAHFNITATAQTALAVAXAATTWVGKKVELNPTKWDVWNKNTGYXTPAVRTMHTL					
	orf29-1						
		ARQWADAHFNITATAQTALSAEAAAGTVWRGKKVELNPTKWDVWNKNTGYXTPAVRHMQL					
		310	320	330	340	350	360
15	orf29a.pep	DGEMAGGNRPFKSITSNSKADASTQPSLQAQLIGEIXXGHAYNKHVIRQQEFTDLNINS					
	orf29-1						
		DGEMAGGNKEIKSLP-NSAAEKRNQNFENSNWSSASFSVHKLTLPNAPGILSPDKVK					
		370	380	390	400	410	420
20		310	320	330	340	350	360

Homology with a predicted ORF from *N.gonorrhoeae*

ORF29 shows 88.8% identity over a 125aa overlap with a predicted ORF (ORF29.ng) from *N.*

gonorrhoeae:

25	orf29.pep	VSPVLPITHERTGFEGVIGYETHFSHGHE	30
	orf29ng	EPGGKYLHFNARGSVKNRVCVAVQTFTDATTAVGPILPITHERTGFEGVIGYETHFSHGHE	102
	orf29.pep	VHSPFDHSDSKSTSDFSGVDGGFTVYQLHRTWSEIHPDEYDGPQAAXYPPPGGARDIY	90
30	orf29ng	VHSPFDHSDSKSTSDFSGVDGGFTVYQLHRTGSEIHPEDYDGPQGGYPPPGGARDIY	162
	orf29.pep	SYVVKGTSTKTKTSIVQAPFSDRWLEENAGAASG	125
	orf29ng	SYHIKGTSTKTKINTVPQAPFSDRWLEENAGAASGFLSRADEAGKLIWENDPKNWRANR	222

The complete length ORF29ng nucleotide sequence <SEQ ID 169> is predicted to encode a protein having amino acid sequence <SEQ ID 170>:

1	MNLFIQKFMM	LFAAAISLLQ	IPISHANGLD	ARLRDDMQAK	HYEPGGKYHL
51	FGNARGSVKN	RVCVAVQTFDA	TAVGPIPLIT	HERTGFEGVI	GYETHFSHGHE
101	HEVHSPFDNH	DSKSTSDFSG	GVDGGFTVYQ	LHRTGSEIHP	EDGYDGPQGG
151	GYPPPGGARD	IYSYHIKGT	TKTKINTVPQ	AFPSDRWLKE	NAGAASGFLS
201	RADEAGKLIW	ENDPDKNWR	NRMDDIRGIV	QGVANPFLTG	FQGLGVGAIAT
251	DSAVSPVTYA	AARKTLQGIH	NLGNLSPEAQ	LAAALQDS	AFAVKDSINS
301	ARQWADAHFN	ITATAQTALA	VIEAATTVMG	GKKVELNPAK	WDVWNKNTGYK
351	KPAARHMQTV	DGEMAGGNKP	LESKNTVTTN	NFFENTGYTE	KVLROASNGD
401	YHGFQSSVDA	FSENGTVIQI	VGGDNTVRHK	LYIPGSGYK	DGNFEYIREA
451	DGKINHRLTV	FNQQLPEK*			

In a second experiment, the following DNA sequence <SEQ ID 171> was identified:

1	atgAATTTCG	CTATTCAAAA	ATTCATGATG	ctgttggcAG	cggcaatatc
51	gatgctgCat	ATCCCCATTA	GTCATGCGAA	CGGTTGGGAT	GCCGTCACAC
101	CGCATGATAT	GCAGGCAAAA	CACTACGAAAC	CGGGTGGCAA	ATACCATCTCG
151	TTTGGTAATG	CTCGCGGCAG	TGTTAAAAAT	CGGGTTGCG	CCGTCCAAAC
201	ATTTGATGCA	ACTGCGGTCG	GCCCCATACT	GCCTATTACA	CACGAACGGA
251	CAGGATTTGA	AGGTGTTATC	GGCTATGAAA	CCCATTTTTT	AGGACACGGA
301	CACGAAGTAC	ACAGTCCGTT	CGATAATCAT	GATTCAAAAA	GCACCTCTGA
351	TTTCAGCGCG	GGCGTAGACG	CGGTTTTCAC	CGTTACCAA	CTTCATCGGA
401	CAGGTCGCGA	AATACATCCC	GCAGACGGAT	ATGACGCGCC	TCAAGGCGGC
451	GGTTATCCGG	AACCACAAGG	GGCAAGGGAT	ATATACAGCT	ACCATATCAA
501	AGGAACCTCA	ACCAAAACAA	AGATAAACAC	TGTTCCGCAAC	CGCCCTTTTT
551	CAGACCGCTG	GCTAAAGAAA	AATGCGCGTG	CGGTTCCGG	TTTTCTCAGC
601	CGTGCAGGATG	AAGCAGGAAA	ACTGATATGG	GAAACAGACC	CGGATAAAAC

5 651 TTGGCGGGCT AACCGTATGG ATGATATTGG CGGCATCGTC CAAGGTGCGG
701 TTAATCCTTT TTAAAGGGG TAGCGATTGG GGCATATTACA
751 CACAGTGC GG TAAGCCCGGT CACAGATACA GCGCTCAGC AGACTCTACA
801 AGGTATTAAAT GATTTAGGAA ATTTAAGTCC GGAAGACAA CTTGCCGCGG
851 CGAGCCTATT ACAGGACAGT GCCTTTGCGG TAAAGACGG CATCAATTCC
901 GCCAGACAAT GGGCTGATGC CCATCCGAAT ATAAACGCAA CAGCCCAAAC
951 TGCCCTTGCC GTAGCAGAGG CCGCAGGTAC GTTTGGCGC GGTAAAAAAG
1001 TAGAACTTAA CCCGACCAAA TGGGATTGGG TTAAAAATAC CGGCTATAAA
1051 AAACCTGCTG CCGGCCATAT CAGAGCTGTA GATGGGAGA TGGCAGGGGG
1101 GAATAGACCG CCTAAATCTA TAACGTCCGA AGGAAAGCT AATGCTGC AA
1151 CCTATCTCTAA GTTGGTTAAAT CAGCTAAATG ACAGAAAATT AAATAACATT
1201 GCGGCTCAAG ATCCAGAGATT GAGTCTAGCT ATTCTAGAGG GTAAAAAATA
1251 TTTTCCAATA GGAAGCTGCAA CTTATGAAGA GGCAGATAGA CTAGGTAAAA
1301 TTTGGGTTGG TGAGGGTGCA AGACAAACTA GTGAGGCGG ATGGTTAAGT
1351 AGAGATGGCA CTCGACAATA TCGGCCACCA ACAGAAAAAA AATCAACATT
1401 TGCAACTACA GGTATTTCAG CAATATTGTA AACTTATACT ATTGATTCAA
1451 ATGAAAAAAG AAATAAAATT AAAAATGGAC ATTTAAATAT TAGGTAA

This encodes a protein having amino acid sequence <SEQ ID 172; ORF29ng-1>:

20 1 MNLPIQKFMML LAAATISMLH IPI SHANGLD ARLRDDMQAK HYPEGGKYHL
5 1 FGNARGSVKN RVCRAVQTFDA TAVGPILPIT HERTGFEVGI GYETHFSG
101 HEVHSPFDNH DSKSTSDFSG GVDGGFTVYQ LHRTGSEIHP ADGYDGPQGG
151 GYPEPQGDARD IYSYHIKGTST TKTINTVPO APFSDRWLKE NAGAASGFLS
201 RADEAGKLIW ENDPKNNWRA NRMDDIRGIV QGAVNPFITG FQGVGIGAIT
25 251 DSAVSPVTD TAAQOTLQGIN DLGNLSPEAQ LAASLLQDS AFAYKDGINS
301 ARQWADAHNP ITATAQTALA VAEAAGTVMR GKVELNPTK WDVKNTGYK
351 KPAARMHMTV DGEAGGNRP PKSITSEKGA NAATYPKLVN QLNEQLNNTI
401 AAQDPRLSLA IHEGKKNFPI GTATYEEADR LGKTVWGEA RQTSGGGWLS
451 RDGTRQYRPP TEKKSQFATT GIQANFETYT IDSNEKRNKI KNGLNIR*

ORF29ng-1 and ORF29-1 show 86.0% identity in 401 aa overlap:

30 orf29ng-1.pep 10 20 30 40 50 60
MNLPIQKFMMLLAAATISMLHIPISHANGLDARLRDDMQAKHYEPGGKYHLFGNARGSVKN
orf29-1 MNLPIQKFMMLFAAATISLLQPISSHANGLDARLRDDMQAKHYEPGGKYHLFGNARGSVKN
35 10 20 30 40 50 60
orf29ng-1.pep 70 80 90 100 110 120
RVCRAVQTFDATAVGPILPITHERTGFEVGI GYETHFSGHGHEVHSPFDNHDSKSTSDFSG
orf29-1 RVYAVQTFDATAVSPVLPIITHERTGFEVGI GYETHFSGHGHEVHSPFDNHDSKSTSDFSG
40 70 80 90 100 110 120
orf29ng-1.pep 130 140 150 160 170 180
GVDGGFTVYQLHRTGSEIHPADGYDGPQGGGYPEPQGGARDIYSYHIKGTSTKTINTVPO
orf29-1 GVDGGFTVYQLHRTGSEIHPEDGYDGPQGGSDYPPPGGARDIYSYVKGSTKTKTINTVPO
45 130 140 150 160 170 180
orf29ng-1.pep 190 200 210 220 230 240
APFSDRWLKENAGAASGFLSRADEAGKLIWENDPKNNWNRNMDDIRGIVQGAVNPFITG
orf29-1 APFSDRWLKENAGAASGFLSRADEAGKLIWESDPKNNWNRNMDDIRGIVQGAVNPFITG
50 190 200 210 220 230 240
orf29ng-1.pep 250 260 270 280 290 300
FQGVGIGAITDSAVSPVTD TAAQOTLQGINDLGNLSPEAQ LAASLLQDS AFAYKDGINS
orf29-1 FQGVGIGAITDSAVSPVTD TAAQOTLQGINDLGNLSPEAQ LAASLLQDS AFAYKDGINS
55 250 260 270 280 290 300
orf29ng-1.pep 310 320 330 340 350 360
ARQWADAHNPITATAQTALAVAEAAGTVMR GKVELNPTKWDVKNTGYKKPAARMHMTV
orf29-1 ARQWADAHNPITATAQTALSAEAAGTVMR GKVELNPTKWDVKNTGYKKPAARMHMTV
60 310 320 330 340 350 360
orf29ng-1.pep 370 380 390 400 410 419

orf29ng-1.pep DGEMAGGNRPKSI-TSEGKANATYPKLVNQLEQNQLNIAAQDPRLSLAIHEGKNKF
| | | | | : | : | : | : | :
orf29-1 DGEMAGGNKPISLSPNSAAEKRRKNFEKFNPNWSSASFDSVHKTLPNPAPGLSPDKVT
 370 380 390 400 410 420

 420 430 440 450 460 470 479
orf29ng-1.pep IGTATYEADRGLKIWVGEGARQTSGGGWLSRDGTQRYPPTKKSQFATTGIOANFETY

orf29-1 RYTSLDGKIITIKNNENNYFRIHDSNRKYPLSDSNNAVTKNLGGQKAQYLQQOQTTH
 430 440 450 460 470 480

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

15 **Example 21**

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 173>:

1 ATGAAAAAAC AAATCACCGC AGCOGTAATG ATGCTGTCTA TGATTGCCCC
51 CGCAATGGCA AACGGCTTGG ACAATCAGGC ATTTGAAGAC CAAATGTTCC
101 ACACGCGGGC AGATGCACCG ATGCAG...

This corresponds to the amino acid sequence <SEQ ID 174; ORF30>:

1 MKKOITAAVM MLSMIAPAMA NGLDNQAFED QMFHTRADAP MQ..

Further work revealed the complete nucleotide sequence <SEQ ID 175>:

1	ATGAAAGGACAA	AAATCACCGC	AGCCAGTAATG	ATGTCGTCTA	TGATTTGCCC
51	CGCAATGGGA	AACCGCTTGG	ACCAATCAGC	TTATTAAGAC	CAGGTTGTCC
101	ACACGCGGGC	AGATGCACGC	ATTCAGTTGG	CGGAGCTTTC	CCAAAGGAGG
151	ATGAAGAGGA	CAGAGGGGGC	GTITCTTCCA	TGTGCTATAT	TGGGTTGTGT
201	TGCCATTGCT	ATTTGGACAC	AGATGATTTT	TAGTTATATG	AGCCGCAAGT
251	RCAGCACTTC	TGTTTAGATG	TGTCTATTGT	TGCGCGATCA	AGGGCGCAAT
301	CTGCTGTGGT	TAGGCGCGCC	ATGAAGAAGT	TTCTTCCTTG	CTAAATATGG
351	ACGTGATGTT	AAATAGGCA	ATTAATATGC	GATATATGCT	TGCGTAAATG
401	ATGATGATGC	ATGATGATGC	ATGATGATGC	ATGATGATGC	ATGATGATGC
451	GTTAATACGC	CAGACACTTT	CTGTCGACAG	GGAAATGGTC	GTCATCGGCC
501	TGTGAAGTCC	BAATCATGCG	ACGATGATAG	AGGACACGCG	TTCCTAAT

This corresponds to the amino acid sequence <SEQ ID 176; ORF30-1>:

1 MKKQITAAVM MLSMIAPAMA NGLDNQAFED QVFHTRADAP MQLAELSQKE
51 MKETEGAFILP LAILGGAIG MWTQHGFSAY TFGRPASVRD VAIAGGLGAI
101 PVGGAAGAGV VSFKAHGREI KIGNNMRIAP FGNRTGHGIP KFPYHRRVT
151 DGTGKTLPGK GIGRHRPWES KSTDRSWKNR F*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF30 shows 97.6% identity over a 42aa overlap with an ORF (ORF30a) from strain A of *N. meningitidis*:

orf30.pep MKKQITAAVVMLSMIAPAMGLNDNQAFEDQM^{10 20 30 40}FHTRADAPMQ
 |||||
 orf30a MKKQITAAVVMLSMIAPAMGLNDNQAFEDQV^{10 20 30 40}FHTRADAPMQLAELSQKEMKKT^{50 60}XGAF⁶⁰LP
 orf30a LXILGGAIGMWTHGFSYAT¹⁰GRPASV²⁰RDVAIAGGLGAT³⁰PGXVGAACKVVSFAKYGRE⁴⁰

70 80 90 100 110 120

The complete length ORF30a nucleotide sequence <SEQ ID 177> is:

5 1 ATGAAAAAAC AATCACCAGC AGCCGTAATG ATGCTGTCTA TGATTGCCCC
 51 CGCAATGGCA AACGCTTTGG ACAATCAGCG ATTTGAAGAC CAAGTGTCTCC
 101 ACACGCGGGC AGATGCGCCG ATGCAGTTGG CGGAGCTTTC TCAAAAGGAG
 151 ATGAAGGANA CAGNCGGGGC GTTTCTTCCA TTGNTATCTC TGGGTGGTGC
 201 TGCCATTGGT ATGTGGACAC AGCATGGTTT TAGTTATGCA ACGACAGGCA
 251 GACCACTTTC TGTTAGAGAT GTTGCTATTG CTGGCGGATT AGGCGCAATT
 301 CCTGGTGTG TAGGCGCCGC AGGAAAGGTT GTTTCCTTTG CTAAATATGG
 351 ACGTGAGATT AAAATCGGCA ATAATATGCG GATAGCCCTC TTCGGTAATA
 401 GAACAGGTCA TCCTATTGGN AAATTTCCCC ATTATCATCG TCGAGTTACG
 451 GATAATACGG GCAAGACTTT GCCTGGACAG GGAATTGGTC GTCATCGCCC
 501 TTGGGAATCA AAATCTACGG ACAGATCATG GAAAACCGC TTCTAA

This encodes a protein having amino acid sequence <SEQ ID 178>:

15 1 MKKQITAAVM MLSMIAPAMA NGLDNQAFED QVFHTRADAP MQLAELSQKE
 51 MKKTXGAFLP LXILGGAATG MWTQHGFSYA TTGRPASVRD VALAGGLGAI
 101 PGKVGAAGKV VSFAYGREI KIGNNMRIAP FGNRTGHPIG KFPYHRRVT
 151 DNTGKTLPGQ GIGRHPWES KSTDRSWKNR F*

ORF30a and ORF30-1 show 97.8% identity in 181 aa overlap:

20 orf30a.pep MKKQITAAVMMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQLAELSQKEMKXTXGAFLP 60
 orf30-1 MKKQITAAVMMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQLAELSQKEMKETEGAFLP 60
 25 orf30a.pep LXILGGAATGMWTQHGFSYATTGRPASVRDVALAGGLGAI PGKVGAAGKVVSFAKYGREI 120
 orf30-1 LXILGGAATGMWTQHGFSYATTGRPASVRDVALAGGLGAI PGKVGAAGKVVSFAKYGREI 120
 orf30a.pep KIGNNMRIAPFGNRTGHPIGKFPYHRRVTDNTGKTLPGQGIGRHPWESKSTDRSWKNR 180
 30 orf30-1 KIGNNMRIAPFGNRTGHPIGKFPYHRRVTDNTGKTLPGQGIGRHPWESKSTDRSWKNR 180
 orf30a.pep FX
 orf30-1 FX

Homology with a predicted ORF from *N.gonorrhoeae*

ORF30 shows 97.6% identity over a 42aa overlap with a predicted ORF (ORF30.ng) from *N. gonorrhoeae*:

40 orf30.pep MKKQITAAVMMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQ 42
 orf30ng MKKQITAAVMMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQLAELSQKEMKETEGAFLP 60

The complete length ORF30ng nucleotide sequence <SEQ ID 179> is

45 1 ATGAAAAAAC AATCACCAGC AGCCGTAATG ATGCTGTCTA TGATCGCCCC
 51 CGCAATGGCA AACGCTTTGG ACAATCAGCG ATTTGAAGAC CAAGTGTCTCC
 101 ACACGCGGGC AGATGCGCCG ATGCAGTTGG CGGAGCTTTC TCAGAAGGAG
 151 ATGAAGGAGA CTGAAGGGGC TTTTCTTCCA TTGCTATCTC TGGGTGGTGC
 201 TGCCATTGGT ATGTGGACAC AGCATGGTTT TAGTTATGCA ACGACAGGCA
 251 GACCACTTTC TGTTAGAGAT GTTGCTGGCG GATTAGCGC AATTCCTGAT
 301 GATGTAGGTG CTGCAGGAAA GGTGTTTCC TTGCTAAAT ATGACGTGA
 351 GATTAAATC GGCATAAATA TCGGGATAG CCCTTTCCGT AATGACACAG
 401 GTCATCCTAT TGGAAAAATT CCCATTATC ATCGTCGAT TACGATAAT
 451 ACGGGCAAGA CTTTGCTTGG ACAGGGAATT GGTCGTCATC GCCCTTGGGA
 501 ATCAAAATCT ACGGACAGAT CATGGAAAA CCGCTTCTAA

This encodes a protein having amino acid sequence <SEQ ID 180>:

55 1 MKKQITAAVM MLSMIAPAMA NGLDNQAFED QVFHTRADAP MQLAELSQKE

51 MKETEGAFLL LAILGGAIG MWTQHGFSYA TTGRPASVRD VAGGLGAIPG
 101 DVGAAGKVVS FAKYGREIKI GNNMRIAPFG NRTGHPIGKF PHYHRRVTDN
 151 TGKTLPGQGI GRHRPWESKS TDRSWKNRF*

ORF30ng and ORF30-1 show 98.3% identity in 181 aa overlap:

5	orf30ng.pep	10	20	30	40	50	60
	orf30-1	10	20	30	40	50	60
10	orf30ng.pep	70	80	90	100	110	
	orf30-1	70	80	90	100	110	120
15	orf30ng.pep	120	130	140	150	160	170
	orf30-1	120	130	140	150	160	170
20	orf30ng.pep	180					
	orf30-1	180					
25	orf30ng.pep						
	orf30-1						

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

30 Example 22

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 181>:

1 ATGAATAAAA CTCCTATCG TGTAATTTTC AACCGCAAC GTGGGGCTGT
 51 GTAGACCGTT GCTGAACTA CCAAGCGCGA AGGTAAAGC TGTGCCGATA
 101 GTGATTCAGG CAGCGCTCAT GTGAAATCTG TTCCTTTTGG TACTACTCAT
 151 GCACCTGTGT GTGCTTCAAA TATCTTTTCT TATTGGGCTT
 201 TTTCTTTATG TTGGCTGTAG GtaccGgCAA TATTGCTTTT GGTGATGGCA
 251 TT..

This corresponds to the amino acid sequence <SEQ ID 182; ORF31>:

1 MNKTLRVIF NRKRGAVXAV AETTKREGKS CADSDSGSAH VKSVFPGTTH
 40 51 APVCSXVNI FSLGFSLCL LAVGTXNIAF ADGI..

Further work revealed a further partial nucleotide sequence <SEQ ID 183>:

1 ATGAATAAAA CTCCTATCG TGTAATTTTC AACCGCAAC GTGGGGCTGT
 51 GGTAGCCGTT GCTGAACTA CCAAGCGCGA AGGTAAAGC TGTGCCGATA
 101 GTGATTCAGG CAGCGCTCAT GTGAAATCTG TTCCTTTTGG TACTACTCAT
 151 GCACCTGTGT GTGCTTCAAA TATCTTTTCT TATTGGGCTT
 201 TTTATGTTTG GCTGTAGGTA CGGCCAATAT TGCTTTTGT GATGGCATT..

This corresponds to the amino acid sequence <SEQ ID 184; ORF31-1>:

1 MNKTLRVIF NRKRGAVVAV AETTKREGKS CADSDSGSAH VKSVFPGTTH
 50 51 APVCSXNIFS FSLGFSLCL AVGTANIAFA DGI..

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.gonorrhoeae*

ORF31 shows 76.2% identity over a 84aa overlap with a predicted ORF (ORF31.ng) from *N.*

gonorrhoeae:

5	orf31.pep	MNKTLYRVIENRRKRGAVXVAEATTKREGKSCADSGSAHVKSVPFGTTHAPVCXVTNIF	60
		: :	
	orf31.ng	MNKTLYRVIENRRKGAVVAVAEATTKREGKSCADSGSGSVYKVSFIPTH-----SKAF	54
10	orf31.pep	SFSLLGFSLCLAVGTXXNIADFAGI	84
	orf31.ng	CFSALGFSLCALGTGVNIAFDAGITDKAAPKTQQAIILOTNGNIGPQVNIQTPTSAGVSV	114

The complete length ORF31ng nucleotide sequence <SEQ ID 185> is:

15	1	ATGACACAAA	CCCTTATGCG	TGTGATTTTC	AACCGCAAAAC	CGGGTGGCTGT
	51	GTGTAGCTGT	GCGTGAAGCA	CAACAGCGGCA	AGGTAAAGCA	TGTCGCCGATA
	101	TGCTGTCGGG	CAGCGGTTAT	TGCAAAATACG	TTTCTCTTAT	TCTACTCAT
	151	TCCTAAGCGCT	TTTGTCTTTC	GTGATTAGCC	TTTTCTTTAT	GTTTGGCTTT
	201	GGGTACGGCT	AATATATGCT	TTGCTGACGG	CATTATTAATC	GATAAAGCTG
20	251	CTCTTAAAC	GAACACAGCC	ACGATCTTCG	AACAGAGTaa	cGCGATACCG
	301	CACGTCAATA	TATCAACCCC	TACITCGGCA	GGGGTTTCTG	TTAATCAATA
	351	TGCCCTAGTT	GTCGTGGGTA	ATTCGCGGGC	GATTTTAAAC	AACAGTCAAG
	401	GCACACACCA	AACACACGTA	GGCGTTTGGG	TTTCAAGGCA	TCTCTGGTTG
	451	ACAAGGGGCG	AAGCAGCTGT	TTGTTTAAAC	CAATATCAAC	CGACGCATCC
25	501	TTTCAACATG	AATGGCTATA	TTGAAGTGGG	TGCAGCGCTG	CGAGAAGTCG
	551	TTATTGCCAA	TCCGCGAGGG	ATTACGATCA	ATGTTGGTGG	TTTTATCAAT
	601	GCTTCCCGTG	CCACTTTGAC	CAGACGGCAA	CCGCAATATC	AAGACGAGGA
	651	CTTTGCGGCG	TTTAAAGATA	GGCAAGACAA	TGCTGTAATC	CGCGGACAGA
	701	GTTTGGATGT	CCGTGATACG	GATTTACAC	GTAATTTGTG	ATGCCACAA
	751	AATCACTCTTG	ACAGTAGCG	CGGAAACAGC	AGGCATTCTG	AA

This encodes a protein having amino acid sequence <SEQ ID 186>:

30

1	MNKTILRVVIF	NRKKRAGVAV	AETTCKREGKS	CADSGSGSVY	VKVSIVFPFTH
51	SKAFCSFSLG	SLKCLALGTI	NIAFADGIIT	DKAAEPCTQQA	TILGTGNGPI
101	QVNIQTPTSA	QSVVNGYQAF	DVGNRGNLR	NRSRNTQTQAL	GGWIGKGNPWL
151	TGRCARVVVN	QINSSEHVVN	NGYIEVGGR	AEVVIANFAG	IADVNGGGINF
201	ASRATLITQG	PQYQAGDFSC	FKIRQGNVAV	AGHGLDARDT	DFTRILVCQQ
251	NHLDQYGRS	RES*			

This gonococcal protein shares 50% identity over a 149aa overlap with the pore-forming hemolysins-like HecA protein from *Erwinia chrysanthemi* (accession number L39897):

40	orf313ng	96	GGNGIPQVNIQPTTSQASVSVNQYVHVDNNGGAILNNSRSN-TQTQGLGWTGQNPNLITRGE	154
	HecA	45	GGNG+P VNI TP +H+G S N+Y F+V NRG LLNN + T +QGG IQ PL R	
45	orf313ng	155	ARVVVVNQINS SHPSQLNGYVGGRRAEVVIANPAGI+VNGGGGFINASRATLTITGQPO+Q	214
	HecA	105	A ++N+K S + S+L GY+EV G+ A VV+ANP G+I +G GF+N R LTTG PQ+Y	
45	orf313ng	215	AAAILNEVVS FPNRSRLAGLEVAGQAANVAVNPYGITCSGGFLNTPRLTLTTGTQPFQD	164
	HecA	165	-ACDGSFGFKRQGNVIAHGILDRATDFE +24 AG SG +R+G +I G GLD+ +24 AAGGLTIGDLGDLIGLGLDASRSDY 193	

Furthermore, ORF31ng and ORF31-1 show 79.5% identity in 83 aa overlap:

[illegible]

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF32 shows 93.8% identity over a 81aa overlap with an ORF (ORF32a) from strain A of *N.*

meningitidis:

		10	20	30	40	50	60
5	orf32.pep	MNTPPFVCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDVDSALRALCPDLPDVV					
	orf32a	MNTPPFSAGXGCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDVDSALRALCPDLPDVX					
		10	20	30	40	50	60
10		70	80				
	orf32.pep	CVHQDIHVRTWHSDAADIDTA					
	orf32a	CVHQDIHVRTWHSDAADIDTAPVXDVIETFACDLPENVLHIIRRHKPLWLKWEYLSAEX					
		70	80	90	100	110	120

15 The complete length ORF32a nucleotide sequence <SEQ ID 191> is:

```

1  ATGAATACTC CTCCTTTTTC TGCTGGANTT TTTTGC AAGG TCATCGACAA
51  TTTGCGGCGAC ATCGGCGGTTT CGTGCGGCTT TGCCCGTGTT TTGCACCGCGG
101 AACTCGGTTG CGAGGTGCAT TTTGGGACGG ACGATGTGTC CGCCTTGCGT
151 GCGCTTTGCC CTGATTTGCC CGATGTTTNC TGCGTTCATC AGGATATTCA
201 TGTCCGCACT TGGCATTCCG ATGGGCGAGA TATTGATACC GCGCTTGCTT
251 NOGATGTGCT CATCGAAACT TTTGCCTGCG ACCTCGCCGA AATGTGCTGT
301 CACATCATCC GCGACACAAA GCGCTTTTGG CTGAANTGGG AATATTTGAG
351 GCGGAGGAGAN AGCAATGAAA GGCTGCACNT GATGCTTCG CCGCAGGAGA
401 GTGTTTCNAAA ATANTTTTGG TTTATGGGTT TCAGCGAANN NAGCGCGCGA
451 CTGATACGCGG AAGCGGATTA CTGCGAAGCC GTCCGTTTGC ATAGCGGAGC
511 TTTGCGCGAAG AGGCTGAGTC TTTCCGCAAA AAGCGNCCC GAATGCTGCG
551 TTTTGGCGCTA TCGSAGCGAT GTTTGGGCAA AGTGGCTGGA AATGTGCGGA
601 CAGCGAGCGA GTCCGTGTAC ACTTTGTGCT GCGGGGCGCG ANATTATCGA
651 CAGCCTCAAA CAAACGCGG TTTATCGCA AGATGCCCTG CAAACGACAG
701 GCGATGTTTT TCAGACGCGA TCCGTCCGCG TCGTCAAAAT CCCTTTCCGTG
751 CCGCAACAGG ACTTCGACAA ACTGCTGCACT TCTGCCGACT GCGCGTCTAT
801 CCGCGCGCAA GACAGTTTGG TGCGCGCCCA GCTTGGCGGC AAACCTTCTC
851 TTTGGGCACAT CTACCGCGAA GATGAGAATG TCCATCTCGA CAAACTCCAT
901 GCCTTTTGGG ATAAGGCACA CGGTTTCTAC ACGCCCGAAA CCGCATCGCG
951 ACACGCGCGC CTTTCAGACG ACCTCAACGG CGGAGAGGCT TTATCCGCAA
1001 CACAACGCGT CGAATGTGG CAAATCTCG AACACATCA AAACGCGCTG
1051 CGGCAAGGCG CGGAGGATTG GAGCCGTTAT CTTTTGGGCG AGCCTCCGCG
1101 ATCCGAAAAA CTCGCGCGCT TTGTTTCAAA GCATCAAAAA ATACGCTAG

```

This encodes a protein having amino acid sequence <SEQ ID 192>:

```

40 1  MNTPPFSAGX FCKVIDNFGD IGVSWRLARV LHRELGWQVH LWTDDVSALR
51  ALCPLDLPDVX CVHQDIHVRT WHSDAADIDT APVDVVIET FACDLPENVL
101 HIIRRHKPLW LKWEYLSAEX SNRLHXHPS PQSVKXFW FMGSEKSGCG
151 LIRERDYCEA VRFDGALRK RIMLPEKNXP EWLFGYRSD VWAKWLEMR
201 QAGSPILTL LL AGAXIIDSLK QNGVIPDAL QNDGDFVQTA SVRIWKIPV
45 251 PQQDFDKLLH LADCAVIRGE DSVFRAQLAG KFFFWHIYVQ DENVHLDKIH
301 APFDKHAHGFY TPETASAHRR LSDDLNGGEA LSAOTRLECW QILQHQHNGW
351 RQGAEDWSRY LFGQPSASEK LAAFVSKHQK IR*

```

ORF32a and ORF32-1 show 93.2% identity in 382 aa overlap:

		10	20	30	40	50	60
50	orf32-1.pep	MNTPPFVCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDVDSALRALCPDLPDVV					
	orf32a	MNTPPFSAGXGCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDVDSALRALCPDLPDVX					
		10	20	30	40	50	60
55		70	80	90	100	110	120
	orf32-1.pep	CVHQDIHVRTWHSDAADIDTAPVDVVIETFACDLPENVLHIIRRHKPLWLKWEYLSAEX					
	orf32a	CVHQDIHVRTWHSDAADIDTAPVXDVIETFACDLPENVLHIIRRHKPLWLKWEYLSAEX					
		70	80	90	100	110	120

60

		130	140	150	160	170	180
	orf32-1.pep	SNERLHLMPS	PQEGVGXKYPWFMGFSEKSGGLIRERDYCEAVRFDTEALRERIMLPEKNAS				
5	orf32a	SNERLHLMPS	PQESVXKKYFWFMGFSEKSGGLIRERDYCEAVRFDGSGALRKRIMLPEKNXP				
		130	140	150	160	170	180
	orf32-1.pep	EWLLFGYRSDV	WAKWLEMRQAGSPMTLLAGTQIIDSLKQSGVIPQDALQNDGDVDFQTA				
10	orf32a	EWLLFGYRSDV	WAKWLEMRQAGSPMTLLAGAXIIDSLKQNGVIPQDALQNDGDVDFQTA				
		190	200	210	220	230	240
	orf32-1.pep	SVRLVKIPFV	PQQDFDQLHLADCAVIRGEDSFVRAQLAGKPFFFWHIYPQDENVHLDKLH				
15	orf32a	SVRLVKIPFV	PQQDFDQLHLADCAVIRGEDSFVRAQLAGKPFFFWHIYPQDENVHLDKLH				
		250	260	270	280	290	300
	orf32-1.pep	AFWDKAHGPHY	PETVSAHRRISDDLNGGEALSATORLECWQTLQHQHNGWRQGAEDWSRY				
20	orf32a	AFWDKAHGPHY	PETVSAHRRISDDLNGGEALSATORLECWQTLQHQHNGWRQGAEDWSRY				
		310	320	330	340	350	360
	orf32-1.pep	LFGQPSAPEK	LAAPVSKHKQIRX				
25	orf32a	LFGQPSAPEK	LAAPVSKHKQIRX				
		370	380				
	orf32-1.pep	LFGQPSASEK	LAAPVSKHKQIRX				
30	orf32a	LFGQPSASEK	LAAPVSKHKQIRX				
		370	380				

Homology with a predicted ORF from *N.gonorrhoeae*

ORF32 shows 95.1% identity over a 82aa overlap with a predicted ORF (ORF32.ng) from *N. gonorrhoeae*:

35	orf32.pep	MNTPPFVCWIFCKVIDNFGDIGVSWRLARVLHRELHGWQVHLWTTDDVSALRALCPDLP	57
	orf32ng	MVMNTYAFVVCWIFCKVIDNFGDIGVSWRLARVLHRELHGWQVHLWTTDDVSALRALCPDLP	60
	orf32.pep	DVPCVHQDIHVRTWHSDAADIDTA	81
40	orf32ng	DVPFVHQDIHVRTWHSDAADIDTAPVPDAVIETFACDLPENVLNIIRRHKLWLNWEYLS	120

An ORF32ng nucleotide sequence <SEQ ID 193> was predicted to encode a protein having amino acid sequence <SEQ ID 194>:

	1	MVMNTYAFV	CWIFCKVIDN	FGDIGVSWRL	ARVLHRELHGW	QVHLWTTDDVS	
45	51	ALRALCPDLP	DVPFVHQDIH	VRTWHSDAAD	IDTAPVPDAV	ISTFACDLEP	
	101	NVLNIIRRHK	PLMLNWEYLS	AEESNERLHL	MPSQEGVGOK	YFWFMGFSEK	
	151	SGGLIRERDY	REAVRFDTEA	LRRLVLPEK	NAPBWLFLGY	RGDVNAKWLMD	
	201	MMWQAGSLMT	LLLAGAQIID	SLKQSGVIPQ	NALNQEAGVF	QTASVRLVXI	
	251	PFVPPQDFDK	LLHLADCAVI	RGEDSFVRTQ	LAKGPFFFWHI	YPQDENVHLMD	
	301	KLHAFWDKAY	GFYTPETASV	HRLLSDDLNG	GEALSATORL	ECGVL*	
50		Further sequencing revealed the following DNA sequence <SEQ ID 195>:					
	1	ATGAATACAT	ACGCTTTTCC	TGCTGTGG	ATTTTTGCA	AGGTCATCGA	
	51	CAATTTCGGC	GACATCGGGC	TTTCGTGGG	GCTCGCCGT	GTTTTCGACC	
	101	GCGAATCGG	TGGCAGGTG	CATTGTGGA	CGGACGACGT	GTCCGCTTGT	
55	151	CGCGCGCTTT	GTCGCGATT	GCCGATGTT	CCCTTGCTC	ATCAGGATAT	
	201	TCATGTCCGC	ACTTGGCATT	CGATGCGGC	AGACATTGAT	ACCGCGCCGC	
	251	TTCCCGATGC	CGTTATCGAA	ACTTTTGCT	CGACCTGCC	CGAAATGTG	
	301	CTGACATCA	TCCGCGGACA	CAAAACGCTT	TGGCTGAATT	GGAAATATT	
	351	CGCGCGGAG	GAAAGCAATG	AAAGGCTGCA	CCTGATGCT	TCGCGCGAGG	
	401	AGGCGCTCA	AAATATTTT	TGGTTATGG	GTTTCAGCGA	AAAAAGCGGC	
60	451	GGTGTGATAC	GCGAACGCGA	TTACCGCGAA	GCCGTCCGTT	TCGATACCGA	

5

10

15 This encodes a protein having amino acid sequence <SEQ ID 196; ORF32ng-1>:

20

```

501 AGCCCTGCGC CGGGGGCTGG TGCTGCCCGA AAAAAACGCC CCCGAATGGC
551 TGCTTTTTCGG CTATCGGGGG GATGTTTGGG CAAAGTGCGT GGACATGTGG
601 CAACAGCGAG GCAGCCTGAT GACCCCTACT CTGGCGGGGG CGCAAAATATg
651 CGACAGCCTC AAACAAGCGC GCGTTATTCG GCAAAAACGCC CTGCAAAAtg
701 aaggcgGTGT CTTTCagacg gcatccgTcC gccttGTCAA AAtcCGGTTc
751 GTGCGcCAAC AGGAcTTCGA CAAATTGCTG CaccctgcCG ACTGCGCCGT
801 GATACGCGCG GAAGACAGTT TCGTGCGTAC CCAGCTTGCC GGA AAAACCCt
851 TTTTTTGGCA CATCTACCCG CAAGACGAGA ATGTCCATCT CGACA AACTC
901 CACGCCCTTT GGGATAAGGC ATACGCCCTT TACACGCCCG AAACCGCATC
951 GGTGCACCGC CTCCTTTTCG ACGACCTCAA CGCGCGAGAG GCTTTATCCG
1001 CAACAACAAC CCTCGAATGT TGGCAAAACC TCACAACAAC TCAAAACCGC
1051 TGGCGGCAAG GCGCGGAGGA TTGGAGCCGT TATCTTTTCC GGCAGCCTC
1101 CGCATCCGAA AACTCGCGC CTTTGTTC AAGCATCAA AAAATACGCT
1151 AG

```

```

1 MNTYAPFVCW IFCKVIDNFG DIGVSWRLAR VLRHRLGWQV HLWTDVDSAL
51 RALCPDLPDV PFVHQDIHVR TWSDAADID TAPVPDAVIE TFACDLPENV
101 LNIIRRHKPL WLNWEYLSAE ESNERLHLMF SPOEGVQKYF WFMGFSEKSG
151 ELIRERDVRE AVRFDTEALR RRLVLPKNA PEWLLFGYRG DVNAKWLMDW
201 QQAGSLMTLL LAGAQIIDL KQSGVIPQNA LQNEGGVQOT ASVRLVKIWF
251 VPQDDFDKLL HLADCAVIRG EDSFVRTQLA GKPFWHIYP QDENVHLDEL
301 HAFWDKAYGF YTPETASVHR LLSDDLNGGE ALSATQRLC WQTLQQHQNG
351 WRQGAEDWSR YLFGQPSASE KLAAFVSKHQ KIR*

```

ORF32ng-1 and ORF32-1 show 93.5% identity in 383 aa overlap:

25

30

35

40

45

50

55

60

65

```

                10      20      30      40      50      59
orf32-1.pep  MNTPPF-VCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDVDSALRALCPDLPDV
orf32ng-1    MNTYAPFVCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDVDSALRALCPDLPDV
                10      20      30      40      50      60

                60      70      80      90      100     110     119
orf32-1.pep  FCVHQDIHVRWTWSDAADIDTAPVPDVVIETFACDLPENVLNIIRRHKPLWLNWEYLSAE
orf32ng-1    PFVHQDIHVRWTWSDAADIDTAPVPDAVIEITFACDLPENVLNIIRRHKPLWLNWEYLSAE
                60      70      80      90      100     110     120

                120     130     140     150     160     170     179
orf32-1.pep  ESNERLHLMFSPOEGVQKYFWFMGFSEKSGGLIRERDYCEAVRFDTEALRRLVLPKNA
orf32ng-1    ESNERLHLMFSPOEGVQKYFWFMGFSEKSGGLIRERDYCEAVRFDTEALRRLVLPKNA
                130     140     150     160     170     180

                180     190     200     210     220     230     239
orf32-1.pep  SEWLLFGYRSDVWAKWLEMMRQAGSPMTLLLAGTQIIDSLSKQSGVIPQDALONDGVFQOT
orf32ng-1    FEWLLFGYRGDVWAKWLDMMQAGSLMTLLLAGAQIIDSLSKQSGVIPQNALQNEGVVQOT
                190     200     210     220     230     240

                240     250     260     270     280     290     299
orf32-1.pep  ASVRLVKIPFVPQDDQLHLADCAVIRGEDSFVRAQLAGKPFWHIYPQDENVHLDEL
orf32ng-1    ASVRLVKIPFVPQDDKLLHLADCAVIRGEDSFVRTQLAGKPFWHIYPQDENVHLDEL
                250     260     270     280     290     300

                300     310     320     330     340     350     359
orf32-1.pep  HAFWDKAGFYTPETVSAHRLSDDLNGEALSATQRLCQWQTLQQHQNGWRQGAEDWSR
orf32ng-1    HAFWDKAYGFYTPETASVHRLSDDLNGEALSATQRLCQWQTLQQHQNGWRQGAEDWSR
                310     320     330     340     350     360

                360     370     380
orf32-1.pep  YLFGQPSAPEKLAAFVSKHQKIRX
orf32ng-1    YLFGQPSASEKLAAFVSKHQKIRX
                370     380

```

On this basis, including the RGD sequence in the gonococcal protein, characteristic of adhesins, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF32-1 (42kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 7A shows the results of affinity purification of the His-fusion protein, and Figure 7B shows the results of expression of the GST-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for ELISA, giving a positive result. These experiments confirm that ORF32-1 is a surface-exposed protein, and that it is a useful immunogen.

Example 24

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 197>:

```

1   ..TTGTTCTCTGC GTGTNAAAGT GGGGCGTTTT TTCAGCAGTC CGGCACAGTCG
51  GTTTCGGGNC AAGACCCCTG TAAATCAGC GGTGTTGCGG CTGTATNCGG
101 ACGAGTGGCG GCA.ACTTCG GTACGTTGGA AAATAGNCGC AACGTGCGAC
151 AGCCTGTGGC TCTGCAAGCT GCTCGGAATG CTGGTGTGCG TATTGTTGCT
201 GCTTTTGGTG CGGCAATATA GGTTCAACTG GGAAAGCAGC CTGTTGAGCA
251 ATGCGCGCTC GGTACGCGCG GTGGAATATG TGGCATGCGT GCGCTCGAAA
301 CTCGGTTTCC CTGTCCCGCA TGCGCGGTGCG GTCATCGAAG GCGCTGTGAA
351 CGGCAATATT GCGGATCGCG GGGCTTGCTG GGGCTGTGCT GTGCGNACGTA
401 TCGCTGCTGA NGGCATCTCG CGCGCGCTG..

```

This corresponds to the amino acid sequence <SEQ ID 198; ORF33>:

```

1   ..LFLRVKVGRF FSSPATWFRX KDPVNQAVLR LYXDEWRXTS VRWKIXATSH
51  SLWLCTLLGM LVSVLLLLLV RQYTFNWEST LLSNAASVRA VEWLAWLPSK
101 LGFFVEDARS VIEGRLENGNI ADARAWSLGL VKSTACXGLL PRL..

```

Further work revealed the complete nucleotide sequence <SEQ ID 199>:

```

1   ATGTTGAATC CATCCGAAA ACTGGTTGAG CTGGTCCGTA TTTTGGACGA
51  AGGCGGTTTT ATTTTCAGCG GCGATCCCGT ACAGCGCAGC GAGGCTTTTC
101 GCCCGCTGGA CGGCAGTACG GAGGAAAAAA TCATCCGCTCG GCGCGAGATG
151 ATTGACAGGA ACGGTATGCT GCGGAGACAG TTGGAACGTG TCGCTGCGGG
201 GTCGTTCTGG TTGTGGGTGG TGGCGGCGAC GTTTGCAATT TTTACCGGTT
251 TTTCACTCAC TTATCTTCTA ATGGACAATC AGGGTCTGAA TTTCTTTTGT
301 GTTTTGGCGG GCGTGTGGGG CATGAATACG CTGATGCTGG CAGTATGGTT
351 GGCAATGTTG TTTCTGCGTG TGAAAGTGGG GCGTTTTCCT AGCAGTCCGG
401 CGACGTGGTT TCGGGGCGAA GACCTGTGAA ATCAGGCGGT GTTGGCGGCT
451 TATGCGGACG AGTGGCGGCA ACCTTCGGTA CTTTGGAAAA TAGGCGCAAC
501 GTCGCACAGC CTGTGGCTCT GCACGCTGCT CGGAATGCTG GTGTGGTATG
551 TGTGTCGCTG TTTGGTGGCG CAATATACGT TCAACTGGGA AAGCAGCGCTG
601 TTGAGCAATG CCGCTTCGGT ACGCGCGGTG GAAATTTTGG CATGGCTGCC
651 GTGCAAACTC GCTTTCCTCG TCCCGGATGC GCGCGCGCTC ATCGAGGCC
701 GCTGAAACGG CAATATTGCC GATGCGCGGG GTTGGTGGGG GTGCTGTGTC
751 GGCAGTATCG CCTGCTACGG CATCTGCCCG CCGCTGCTGG CTTCGGTAGT
801 GTGTAAATCT CTTTTGAAAA CAAGCGAAAA CCGATTGGAT TTGGAANAAGC
851 CCTATTATCA GCGGCTCATC CGCCGCTGGC AGAACAATAT CACCGATGCG
901 GATACGCGCT GGGAAACCGT GTCCGCGGTT TCACCGAAAA TCATCTTTGAA
951 CGATGCGCGG AATATGGCGG TCATGCTGGA GACCGAGTGG CAGGACGGCG
1001 AATGGTTTGA GGGCAGGCTG GCGCAGGAAT GCGTGTGATA GGGCGTGTCC
1051 ACCAATCGGG AACAGGTGCG CGCGCTGGAG ACAGAGCTGA AGCAGAAACC
1101 GCGGCAACTG CTTATCGCGG TCGCGGCCCA AACTGTGCCG GACCGCGCGG
1151 TGTTCGGGCA GATTGTCCGA CTCTCGGAAG CCGCGCAGGG CCGCGCGGTC

```

1201 GTGCAGCTTT TGGCGGAACA GGGGCTTTCA GACGACCTTT CGGAAAAGCT
1251 GGAACATTGG CGTAACGCGC TGGCCGAATG CGGCGCGGCG TGGCTTGAGC
1301 CTGACAGGGC GGCGCAGGAA GGGCGTTTGA AAGACCAATA A

This corresponds to the amino acid sequence <SEQ ID 200; ORF33-1>:

5	1	MLNPSKSLIVE	LVRLEDEGFF	IFSGDQVATAT	EALRRVDGST	EKKIIRRAEM
	51	IDNRNMLRET	LVRVRAGSFF	LDLNUVATFAF	FTGRRSYTLL	MDNQLNLFLL
	101	VLAGVLGMNT	MLNLVWLAMI	FLRWKVGRRF	SSPATWFRGK	DPVNGAVLKIL
	151	VEAGVQRQPS	RWKIGATSHS	LWLCITLIGM	VSLLILLMLL	TYTNNFWSLT
	201	LSNAAGSVRA	MLLAWLPKSL	GLFFVDFARV	TEKRLNGNTA	DARWSLSWLL
	251	SIACGYCLYF	RLNLVWVLCI	LKPTSGENGL	HEKRYGVAVI	RWKQKQVWLL
	301	FTREYRERET	PEKLLADAF	KWYVWVWV	QDGFQVGL	AGENLDKGVV
	351	TREYQVALE	FTLKQKPAEW	LILVGRACVT	DGRVLRIORV	LSFAAQGGAV
	401	VOILAFOGLS	PDLSKELEBW	RNALAEAGAA	WLEPDRAAQE	GRKLDQ*

Computer analysis of this amino acid sequence gave the following results:

15 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF33 shows 90.9% identity over a 143aa overlap with an ORF (ORF33a) from strain A of *N.*

meningitidis:

[illegible]

The complete length ORF33a nucleotide sequence <SEQ ID 201> is:

40	1	ATGTTGGATC	ATCCCGGAA	ACGTTGTTCG	CTGTGGCCGA	TTTTGGGAAG
	51	AGCGCGCTTT	ATTTCCTGAG	CGCATTCAGG	CGACGAGCGA	GAGCCTTTGC
	101	CGCGCGCTGGA	CGGCAATACG	GAGGAAAGCA	TCATCCGTCG	GGCGAAGATG
	151	ATGACACAGA	ACGATGATCG	CGGGGAGACG	TGTGAACCTG	TGCGTCGGGG
45	201	GTCGCTCTGG	TGTGGGTGGG	CGGGCGCGAC	TGTTGCGGTT	NTTACCGNIT
	251	TTTCAGTAC	TATATCTTCA	TAGACAAATC	AGGGTCTGAA	TTTCTTGT
	301	TTTCTGGCG	TCGCTGNTGG	CGAATAATCG	TCGTATGCTG	CAGTATGTTT
	351	GGCAATGTGT	TGCTTSCGGG	TGAAGAGTGG	CGGTTTTTTC	AGACGTCGGT
50	401	CGACGCTGCT	GTACGGCAAA	ATGCGGCTGG	GTTCGAGGCG	TTGCGGCTGG
	451	TATCGGCTG	AGTGGGCGN	ACCTTCTGCT	CTGTGGAAAG	TAGCGCGAAC
	501	GTGCGACAGC	CTTGCGCTCT	CGACGCTGCT	CGGAATGCTG	TCTGCGATAT
	551	TGTTGCTGCT	TTTGGTGGCG	CAATATACGT	TCACATCTGG	AAGCACGCTG
55	601	TTGGGCGATT	CGTCTTCGGT	AGCGCTGGTG	GAAATGTGGA	CATGGCTGCC
	651	TGCGAAACTG	GATTTTCCCG	TGCGTATCG	CGGGCGGCTG	ATCGAAGGTC
	701	GTCTGAACGG	CGAATATCCG	CGTCCGCGGG	CTGCTCGGG	GTCTCGTGTC
	751	GGCAATATCG	CTTCTGACGG	CACTCTCGCG	CGCGCTTGG	CTTGGGCGTG
60	801	ATGCAARAAT	CCNTTGNAAA	CAAGCGAAGA	CGGCTTGGAT	TTGGAAGAAAG
	851	NNNNNNNTT	NNNNGTCAT	CGCGCTCGTG	AGACAAAAAT	CACCGATGCG
	901	GATCGACGCG	GGGACACGCT	TTGCGGCAAG	TTGCGGCAAG	TTGCGGCAAG
	951	CCATGCGCCG	AAAATCGCGG	TCATCTCGGA	ACCGCAATAG	CAGGACAGCG
60	1001	AATGGTTTCA	GGGACAGGCT	CGCGCAGGAAT	CGGCTGGATA	GGGCGTTCGC
	1051	GCCAAATCCG	AGACAGGTGC	CGCGCTGGAG	ACGACAGCTGA	AGACAAAGAC

1101 GGCGCAACTS CTTATCGGCG TGGCGGCCCA AACTGTGCCC GACCGCGGGG
 1151 TGTCGCGGCA GATCGTCCGA CTTTCGGAAG CGGCGCAGGG CGGCGCGGTG
 1201 GTGCANCTTT TGGCGGAACA GGGGCTTTCA GACGACCTTT CGGAAAAGCT
 1251 GGAACTTTGG CGTAACGCGC TGACCGAATG CGGCGCGGCG TGGCTGGAAC
 1301 CCGACAGAGC GGGCGAGGAA GGCCTCTGA AAACCAAGC CGGCACTTGA

This encodes a protein having amino acid sequence <SEQ ID 202>:

1 MLNPSRKIVE LVRILEEGGF IFSGDPVQAT EALRRVDGST EEKIIRRAKM
 51 IDNRNMLRET LERVAGSFW LWVAATFAF XTXFVITYLL MDNQGNLFFL
 101 VLAGVXGMNT LMLAVWLAML FLRVKVGRRF SSPATWFRGK DPNVQAVLRL
 151 YADEWRXPSV RWKIGATSHS LWLCTLLGML VSVLLLLLRV QYTFNWESTL
 201 LGDSSSVRLV EMLAWLPKIL GFPVDPARAV TEGRLNGNIA DARAWSGLLV
 251 GSIACYGILP RLLAWAVCKI LXXTSENGLD LEKXXXXXXI RRWQKITDA
 301 DTRRETSAVS SPKIVLNDAP KWAVMLETEW QDGEWFEGR LACEWLDKGA
 351 ANREQVAALE TELKQKPAQL LIGVRAQTVP DRGVLRQIVR LSEAAQGSVA
 401 VXLLAEQGLS DDLSEKLEHW RNALTECGAA WLEPDRAAQE GLRKTNDRT*

ORF33a and ORF33-1 show 94.1% identity in 444 aa overlap:

		10	20	30	40	50	60
orf33a.pep		MLNPSRKIV	ELVRILEEGGF	IFSGDPVQAT	EALRRVDGST	EEKIIRRAKM	IDNRNMLRET
orf33-1		MLNPSRKIV	ELVRILEEGGF	IFSGDPVQAT	EALRRVDGST	EEKIIRRAKM	IDNRNMLRET
		10	20	30	40	50	60
orf33a.pep		70	80	90	100	110	120
orf33-1		70	80	90	100	110	120
		70	80	90	100	110	120
orf33a.pep		130	140	150	160	170	180
orf33-1		130	140	150	160	170	180
		130	140	150	160	170	180
orf33a.pep		190	200	210	220	230	240
orf33-1		190	200	210	220	230	240
		190	200	210	220	230	240
orf33a.pep		250	260	270	280	290	300
orf33-1		250	260	270	280	290	300
		250	260	270	280	290	300
orf33a.pep		310	320	330	340	350	360
orf33-1		310	320	330	340	350	360
		310	320	330	340	350	360
orf33a.pep		370	380	390	400	410	420
orf33-1		370	380	390	400	410	420
		370	380	390	400	410	420
orf33a.pep		430	440	450			
orf33-1		430	440	450			
		430	440	450			

Homology with a predicted ORF from *N.gonorrhoeae*

ORF33 shows 91.6% identity over a 143aa overlap with a predicted ORF (ORF33.ng) from *N. gonorrhoeae*:

5	orf33.pep	LFLRVKVGRRFSSPATWFRXKDPVNQAVLR	30
	orf33.ng	LMNDQGLNFFLVLAGVLMNTLMLAVLWATLFLRVKVGRRFSSPATWFRGKGPVNQAVLR	100
10	orf33.pep	LYXDWRXTSVRWKIXATSHSLWLCTLLGMLVSVLLLLLVQRQYTFNWESTLLSNAASVRA	90
	orf33.ng	LYADQWRQPSVRWKIGATAHSLWLCTLLGMLVSVLLLLLVQRQYTFNWESTLLSNAASVRA	160
	orf33.pep	VEMLAWLPSPKLGFFVPDARSVIEGRINGNIADARANSGLLVXSIACKGILPRL	143
	orf33.ng	VEMLAWLPSPKLGFFVPDARAVIEGRINGNIADARANSGLLVGSIVCYGILPRLAWVCK	220

- 15 An ORF33ng nucleotide sequence <SEQ ID 203> was predicted to encode a protein having amino acid sequence <SEQ ID 204>:

20	1	MIDRDMRLRD	TLERVAGSF	KLWVVVASM	FTAGFSGTYL	LMNDQGLNFF
	51	LVLAVGLMNT	TLMLAVLWAT	LFLRVKVGRR	FSSPATWFRG	KGPVNQAVLR
	101	LYADQWRQPS	VRWKIGATAH	SLWLCTLLGM	LVSLLLLLV	ROYTFNWEST
	151	LLSNAASVRA	VEMLAWLPSP	KGFFVPDARA	VIEGRINGNI	ADARANSGLL
	201	VGSIVCYGIL	PRLAWVCK	ILLKTSNGL	DLEKTYQAV	IRRWNKQITD
	251	ADTRRETVSA	VSPKIVLND	PKWALMLETE	WQDQWFEGR	LAQEWLDKGV
	301	AANREQVAAL	ETELKQKPAQ	LLIGVRAQTV	PDRGVLRQIV	RLSEAAQGGGA
	351	VVQLLAEGQL	SDDLSEKLEH	WRNALTECGA	AWLEPDRVAQ	EGRLKQD*

- 25 Further sequence analysis revealed the following DNA sequence <SEQ ID 205>:

30	1	ATGTTGaatC	CATCCCGaAA	ACTGgttgag	ctGgTCCgtA	Ttttgaataa
	51	agggggTtTT	atttttcagcg	cgcatcctgt	gcagggcgacg	gaggctttgc
	101	gccgcgttgg	cgccAGTACG	GaggAaaaaa	tcttcctgtcg	GGCGGAGAtg
	151	atcgACAGGG	accgtatggt	gogggACaCg	TgggaacGTG	TGCGTGCGgg
	201	gtcgtTctcg	TTATGggTGG	TggTggCAtC	gATGATGtT	aCGCGCGAT
	251	TTTCAGgac	ttatctctTG	ATGGAaAtC	AGGGGctGAA	TcTCTTTTAA
35	301	GTTTtggcgG	GAGTGTgggG	CATGaatacG	ctgATGCTGG	CAGTATGgtt
	351	gCAACGCTTG	TTCCTGCGCG	TGAAAGTGGG	ACGgTtTTTC	AGCAGTCCGG
	401	CGACGTGGTT	TCGGGGCAAA	GGCCCTGTAA	ATCAGGCGGT	GTTGCGGCTG
	451	TATGCGGACC	AGTGGCGGCA	ACCTTCGGTA	CGATGGGAAA	TAGGCGCAAC
	501	GGCGCACAGC	TTGTGGCTCT	GCACGCTGCT	CGGAATGCTG	GTGTCGGTAT
	551	TGCTGCTGCT	TTTGGTGGCG	CAATATACGT	TCAACTGGGA	AAGCACGCTG
40	601	TTGAGCAATG	CCGCTTCGGT	ACGCGCGGTG	GAAATGTGG	CATGGCTGCC
	651	GTCGAAACTC	GGTTCCTCGT	TCCCGATGTC	GCGGGCGGTC	ATCGAAGGTC
	701	GCTGGAACGG	CAATATTGCC	GATCGCGGGG	CTTGCTGGCG	GCTGCTGGTC
	751	GGCAGTATCG	TCTGCTACGG	CATCCTGGCG	CGCCTCTTGG	CTTGGGTAGT
	801	GTGTAAATATC	CTTTTGAAAA	CAAGCGAAAA	CGgattgGAT	TTTGAAAAAA
	851	CCTATTATCA	GGCGGTATC	CGCGGTGCG	AGAACAATA	CACCGATGCG
45	901	GATACGCTC	GGGAACGCT	GTCGCGCTG	TGCGCGAARA	TGCTCTTGA
	951	CGATCGCGCG	AATGGGCGCG	TCATGCTGCA	GACCGAGTGG	CAGCAGCGCC
	1001	AATGGTTCGA	GCGCAGGCTG	GCGCAGGAAT	GGCTGATATA	GGCGCTTGCC
	1051	GCCAATCGSG	AACAGGTTGC	CGCGCTGGAG	ACAGAGCTGA	AGCAGAAACC
	1101	GGCGCAACTG	CTTATCGGCG	TACGCGCCCA	AACTGTGCCG	GACCGGGGGG
	1151	TGCTGCGGCA	GATTGTGCGG	CTTTCGGAG	CGGCGCAGGG	CGCGCGGGTG
50	1201	GTGCAGCTTT	TGGCGGAACA	GGGGCTTTCA	GACGACCTTT	CGGAAAAGCT
	1251	GGAAACATTGG	CGTAACGCGC	TGACCGGAATG	GCGCGCGCGC	TGGCTTGAGC
	1301	CTGACAGGGT	GGCGCAGGAA	GGCCGTTTGA	AAGACCAATA	A

This encodes a protein having amino acid sequence <SEQ ID 206; ORF33ng-1>:

55	1	MLNPSRKIVE	LVRILNKGFF	IFSGDPVQAT	EALRRVDGST	EEKIFRAEM
	51	IDRDMRLDRT	TLERVAGSFW	LWVVVASMFF	TAGFSGTYLL	MDNQGLNFFL
	101	VLAVGLMNT	LMLAVLWATL	FLRVKVGRRF	SSPATWFRGK	GPVNQAVLRL
	151	YADQWRQPSV	RWKIGATAHS	LWLCTLLGML	VSLLLLLVLR	QYTFNNESTL
	201	LSNAASVRAV	EMLAWLPSPK	GFFVPDARAV	IEGRINGNIA	DARANSGLLV

```

251 GSIVCYGILP RLLANVVCKI LLKTSENGLD LEKTYQAVI RRWONKITDA
301 DTRRETVSAV SPKIVLNDAP KWALMLETEW QDGQWFEGR LAQEWLDKGVA
351 ANREQVAALE TELKQKPAQL LGVRAQTVP DRGVLRLQIVR LSEAAQGGAV
401 VQLLAEQGLS DDLSEKLEHW RNALTECGAA WLEPDRVAQE GRLLKQD*

```

5 ORF33ng-1 and ORF33-1 show 94.6% identity in 446 aa overlap:

		10	20	30	40	50	60
orf33-1.pep		MLNPSRKLVELVRILDEGGFIFSGDPVQATEALRRVDGSTEELIRRAEMIDRNRMLRET					
orf33ng-1		MLNPSRKLVELVRILNKGGFIFSGDPVQATEALRRVDGSTEELIRRAEMIDRNRMLRET					
		10	20	30	40	50	60
orf33-1.pep		LERVRAGSFVLWVVAATFAFFTGFVSVTYLLMDNQGLNFFLVLAGVLGMNTIMLAVLML					
orf33ng-1		LERVRAGSFVLWVVAATFAFFTGFVSVTYLLMDNQGLNFFLVLAGVLGMNTIMLAVLML					
		70	80	90	100	110	120
orf33-1.pep		LERVRAGSFVLWVVAATFAFFTGFVSVTYLLMDNQGLNFFLVLAGVLGMNTIMLAVLML					
orf33ng-1		LERVRAGSFVLWVVAATFAFFTGFVSVTYLLMDNQGLNFFLVLAGVLGMNTIMLAVLML					
		70	80	90	100	110	120
orf33-1.pep		FLRVKVRGRFSSPATWFRGKDPVQAVRLRYADEWRQPSVVRWKIGATSHSLWLCTLLGML					
orf33ng-1		FLRVKVRGRFSSPATWFRGKDPVQAVRLRYADEWRQPSVVRWKIGATSHSLWLCTLLGML					
		130	140	150	160	170	180
orf33-1.pep		FLRVKVRGRFSSPATWFRGKDPVQAVRLRYADEWRQPSVVRWKIGATSHSLWLCTLLGML					
orf33ng-1		FLRVKVRGRFSSPATWFRGKDPVQAVRLRYADEWRQPSVVRWKIGATSHSLWLCTLLGML					
		130	140	150	160	170	180
orf33-1.pep		VSVLLLLLRQYTFENWESTLLSNAASVRAVEMLAWLPSKLGFPVPDARAVIEGRNLGNIA					
orf33ng-1		VSVLLLLLRQYTFENWESTLLSNAASVRAVEMLAWLPSKLGFPVPDARAVIEGRNLGNIA					
		190	200	210	220	230	240
orf33-1.pep		VSVLLLLLRQYTFENWESTLLSNAASVRAVEMLAWLPSKLGFPVPDARAVIEGRNLGNIA					
orf33ng-1		VSVLLLLLRQYTFENWESTLLSNAASVRAVEMLAWLPSKLGFPVPDARAVIEGRNLGNIA					
		190	200	210	220	230	240
orf33-1.pep		DARAWSGLLVGSIACYGILPRLANVVCKILLKTSENGLDLEKTYQAVIRRWONKITDA					
orf33ng-1		DARAWSGLLVGSIACYGILPRLANVVCKILLKTSENGLDLEKTYQAVIRRWONKITDA					
		250	260	270	280	290	300
orf33-1.pep		DARAWSGLLVGSIACYGILPRLANVVCKILLKTSENGLDLEKTYQAVIRRWONKITDA					
orf33ng-1		DARAWSGLLVGSIACYGILPRLANVVCKILLKTSENGLDLEKTYQAVIRRWONKITDA					
		250	260	270	280	290	300
orf33-1.pep		DTRRETAVSPKIIILNDAPKWAVMLETEWQDGQWFEGR LAQEWLDKGVAANREQVAALE					
orf33ng-1		DTRRETAVSPKIIILNDAPKWAVMLETEWQDGQWFEGR LAQEWLDKGVAANREQVAALE					
		310	320	330	340	350	360
orf33-1.pep		DTRRETAVSPKIIILNDAPKWAVMLETEWQDGQWFEGR LAQEWLDKGVAANREQVAALE					
orf33ng-1		DTRRETAVSPKIIILNDAPKWAVMLETEWQDGQWFEGR LAQEWLDKGVAANREQVAALE					
		310	320	330	340	350	360
orf33-1.pep		TELKQKPAQLLIGVRAQTVPDRGVLRLQIVRLSEAAQGGAVVQLLAEQGLSDDLSEKLEHW					
orf33ng-1		TELKQKPAQLLIGVRAQTVPDRGVLRLQIVRLSEAAQGGAVVQLLAEQGLSDDLSEKLEHW					
		370	380	390	400	410	420
orf33-1.pep		TELKQKPAQLLIGVRAQTVPDRGVLRLQIVRLSEAAQGGAVVQLLAEQGLSDDLSEKLEHW					
orf33ng-1		TELKQKPAQLLIGVRAQTVPDRGVLRLQIVRLSEAAQGGAVVQLLAEQGLSDDLSEKLEHW					
		370	380	390	400	410	420
orf33-1.pep		RNALTECGAAWLEPDRVAQEGRLKQDQX					
orf33ng-1		RNALTECGAAWLEPDRVAQEGRLKQDQX					
		430	440				
orf33-1.pep		RNALTECGAAWLEPDRVAQEGRLKQDQX					
orf33ng-1		RNALTECGAAWLEPDRVAQEGRLKQDQX					
		430	440				

Based on the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 25

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 207>:

```

1 ..CAGAAGATT TGTCGAGAAT TCCTTTATGG GGTGTTGGCG CCGTGTTTTT

```

5 51 CGGGGTGTCC GGTCTGGTAT GGTTTTCTTT GGGCGTTTCT TT..AGTGGCG
101 CCGCTTTTTC GCGTGTATCT TTTCCGGGTT CGGACGGGG GACGTTTGTG
151 GGCAGTACGG GGGTTTCTTT GAGTGTGTT TCAGCTTGTG TTCC..GGCGT
201 CGTCCGGCTG CTTGTCCGTT TGACGTGTGT CGGACAGTTG C.. GTTTGA
251 CCGGCTTTT CTTGGGTCGG CAGAGGGAGC TCATCTCTCT GCCGTTTTCG
301 TCTGTGCCGT CCGGCTGTGC GGGTTCGGAT GAGGCGGCGT GGTGGTGTTC
351 GGGTTGGGCG GCATCTGTT CCGACTACGC GTTTGGCAG CCAGAATTCCG
401 GTTTCGGGG GGCCTGCGGT GTGTGCGGT TCGGCTTGAA GGGTTTGTTC
451 GTCC..

10 This corresponds to the amino acid sequence <SEQ ID 208; ORF34>:

1 ..QKSLSRISLW GLGGVFFGVG GLVWFSLVGS XECACFSGVG FRSGRGRTFV
51 GSTGVSLSVF SACVXGVVRL PVGLSCVGR LXXLIRFFLGA AGDVILLPLS
101 SVPSGCAGSD EAAWCSGWA ASCPTTFFGS QNSVSRGLSV CCGSA*RVLS
151 S..

15 Further work revealed the complete nucleotide sequence <SEQ ID 209>:

1 ATGATGATGC CGTTCATAAT GCTTCCTTGG ATTGCKGGTG TGCCTGCCGT
51 GCGGGGTGAG AATAGGTTGT CAGAATTTC TTTATGGGT TTGGCGCGCG
101 TGTTTTTTCG GGTGTCCGCT TTGCTATAGT TATCTTTGG CTTTCTTGG
151 GGCTGCGGCT GTTTCTCGGG TGTTTCTTTT CCGGGTTTGG GACGGGGAGC
201 GTTTTGCGGC AGTACGGGGG TTTCTTTGAG TGTGTTTCA GCTGTGTTTC
251 CCGGCTGCTC CGGCTCCGCT TCGGTTTGA CTTGTGTCGG AGGTTGCCGT
301 TTGACCCGGT TTTTCTTGGG TCGGCGAGGG GACGGCAGTC CGCTGCCGCT
351 TTCGCTGTGT CCGTCCGGCT GTGCGGGTTC GGATGAGGCG CGCTGTGTGT
401 GTTCGGGTTG GCGGCGCATC TGTCCGACTA CCGCGTTTGG CAGCCAGAA
451 TCGGTTTTCG GGGGGCTGTC GGTGTGTTGC GGTTCGGCTT GAAGGGTTTT
501 GTGCGCGTTC GGGTTGAATG TGCTGACGAT GCCTATTGCC AATGCGCCGA
551 TGCGCGGCTC ACAGATGAGC AATACGGCGC GTATCAGGAG TTTGGGGGTC
601 AGCCTGAAGG GTTGTTCGGG TTTTTTGGCC ATTTTGAITG TGCTTTTGGG
651 GTGTCGGGCA ATGCGGCTCG AAGGCGGTTT AGACGGCATT CCGGAGTCAG
701 CGTTGACGCT AGTTTGGTGA GAGGCGGATG ACCTTTTGTA CGCGACGCT
751 GGTGCTGACT TTTTGGGTAA TCTGCGGCTT TCTTTCGGGG GTGAGGATGC
801 CCATAACGTA GGTTAGCTTG CCGTAGGTAA CGATTTTGAC CGCGGCTGT
851 GTGGCGGGCG TGATGCCCAA CAGCGTGGCG CGGACTTTGG ATGTGTTCCA
901 AGTGTCGCGG CGGATGTCGC CGGCGGTGCG CGGACGGAG CGGACGATG
951 TATAGTTGTA CCGCGCTTCG CGGCGCTGTT CGGAACCTGC AATCTGACCC
1001 ACGAACGTTT TTTCCGCTTC GGTGGCGACT TCTCCGAGCA GCAGCAGCTG
1051 GCGGTTGTAG CGACGACGCG AGATTTGGGG CGTGATGCTT TTGTTTGTGT
1101 TGTTTTGGCG CAGATAGGAA CCGGCGGTGG TTTGATAGC CAACGCCATA
1151 ACGTTGTCGT CGGTTTGGCG CGCGGTGGTT CGGCGGTGCA CGGCGGATTT
1201 CGCGCCGACG CGGCGGCTTC CGATTACTGC GCTGACGAG CGGCTAAGGG
1251 CAAGGCTGAA AATGCGCGCA ATCAGGTCGC GGACGCTGTG CGGTTTGGGT
1301 TTCATCGGCT GCTTCCCTTC TTGGGCGTTT CAGACGGCAT TGCTTTGGCG
1351 CATGCCGTCT GA

This corresponds to the amino acid sequence <SEQ ID 210; ORF34-1>:

45 1 MMMPFIMLPW IAGVPAVPGQ NRLSRISLWG LGGVFFGVSG LVWFSLVGSL
51 GCACFSGVSF RSGRGRTFVG STGVSLSVFS ACVPASSGCL SV*AVSAGCG
101 LTRFLFLGAG DGSPLPLSSV PSGCAGSDEA AWCSCWAAS CPTTFFSGN
151 SVSRGLSVLC GSA*RVLPSE GLANLWMPJA NAFPMALQMS NTAIRSLVZ
201 SLKGLFCFFA ILIVLLCGRA MPESGGSDCI AESALDVVLV EQDDFLYADG
251 CADFLGNRL PFGGSDAHNV GYVAVGNDFD ARLCGADGVA QRGADPGCVF
301 SVAGDVAGSA RQGGDGNIVV HAFGGLFGTC NLITDELFFAF GGDLSQQQOV
351 AVVADDDBLG RVAFGLVLVA QIGTCGGEDT ORHNVVVGLR AGGSVADGGEF
401 RADGASDYC ADAAAKGKAE NGGNQAGDGV RFGHRVLVLF LGVSDGLIALR
451 HAV*

55 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF34 shows 73.3% identity over a 161aa overlap with an ORF (ORF34a) from strain A of *N. meningitidis*:


```

                    10         20         30
orf34.pep          QKLSRSLMLGLGVFFGVSGLVWFSLGVXS-----CAC
                   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
orf34a             MMXFXIMLPWIAIGVPFVGQKLSRLXSLMGLGGXFFGVSGLVWFSLGVXSGLVGSXGCAG
                   10         20         30         40         50         60
5                  40         50         60         70         80         90
orf34.pep          FSGVSRFGSGRGYFVGSTGVSLSVFSACVXGVVRLEPVGLSCVCRLLX-----LTRFLFGA
                   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
10 orf34a           FSGVSRFGSGRGYFVGSTGVSLSVFSAC-----PASSGCLSVXAVSAGCGLTRXFAGA
                   70         80         90        100        110
                          100        120        130        140        150
orf34.pep          AGDVILLPLISVSPGCGASDEAAWCSGWAASCPTTTPFGSQNSVSRGLSVCCGSAXRVLS
                   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
15 orf34a           AGDGSPLPLISVSPGCGAGADEEAXXC SGWAASCPTTTPFGSQNSVSRGLSVCCGSVWRVLS
                   120        130        140        150        160        170
20 orf34.pep       S
orf34a             PFGXNVLTMPIANAPMAVIQMSNTARISLGVSLKGLFXFFAILIIVLLGRAMPSEGGSD
                   180        190        200        210        220        230

```

The complete length ORF34a nucleotide sequence <SEQ ID 211> is:

25	1	ATGATGATGNC	CTGTCTNATAAT	CGCTCACTTGG	ATTGCGGGGT	TGCGTCGCCGT
	51	CGGGGTCAG	AGAGGATGTT	GAGTATGGGT	TAGAGCGGCT	
	101	TGTTTTTCCG	GGTCTCCGTT	TGTGATATGTT	TTCCTTTGGG	CGCTTCTTNT
	151	TCTTTGGGTG	TTCTTCTGGG	CTRGAGCTGT	TTTTCCGGGT	TCTTCTTTTG
	201	GGGTCGGGA	CGGGGGACGT	TGTTGGGACG	TACNGGGTTT	TCTTTTGAAGT
30	251	TGTTTTCAGC	TTTGTCCTGG	GGTTCGCTCC	GGTCGGCTGT	TCTTTNAGCT
	301	GTGTGGGACG	TTGCGGGTTT	AGCCCGGNTT	TTCTTNGTGT	CGCGAGGGGA
	351	CGGACGTCCG	CTGCGCGGTT	GATCTGTGCC	TGCGCGCTGT	CGGGTCCGG
	401	ATAGAGAGCG	CTGCTNGTGT	TCCGGTTTGG	GGGATCTGT	TCCGACTACG
	451	CCGTTTGGCA	SCNAGAAATC	GGTTTCCGGG	CGGCTGTCCG	TGTTTTCGGG
35	501	TCTGNTTNG	AGSGTTTGT	CMCGGTTCCG	TGNGAATGTT	CTGACGATCT
	551	TTATGGCCAA	TGCGCGTACG	CGGCTGATAC	AGAGAGGCTA	TACGAGGCTA
	601	ATCGAGTATG	CGGGTCTGAG	CTGCTTTGTT	TCTTTTGGCT	TCTTTTGGCT
	651	TTTGATTTGG	TTCTTGGGCT	TGCGGCAATG	CGCCTCTGAA	CGCGCTTCAG
	701	ACGSCATGTC	CGAGTACCGG	TTGAGACTAG	TTTTNGTAGA	GGGTTGATAG
40	751	TTTTTTCAC	CGACCGGTGG	TGCTTGACTTT	TTGGGTATAT	TGCGCCTGTT
	801	CTTCGGGGGT	GAGGATGCCA	TGAGCTAGG	TTAGCTTGCC	TATAGATAAG
	851	ATTTTGAAGC	GGGCTCTGTG	GGCGGGGCTG	ATGCCCAACA	CGGTGGCGCG
	901	GACTTTGGAT	GTGTTCCAGT	TGTCCGCGCG	GATGTCCGCG	CGAGTCCGCG
	951	CGAGGGAGCA	GACGGTAAATG	TANTTTGTCA	CGCCTTCGCG	GGCCTCTTCC
45	1001	GAACTGCA	CTCAGCAGC	GAAGTCTTTC	TGCGCTTCGG	TGGGACATCT
	1051	TCCGACGACG	AGCAGGTGGC	GGTTGTAGCA	GACAAAGGAG	ATTTGGGGGG
	1101	TGNTACCTTT	GAGTTTGGTG	TTTTTGGCCA	GATAGAGGAG	CGGCTGGTGT
	1151	TGCAATACCA	CGCGCTATAC	GTGTCTCTCG	CTTNGCCGCG	CGGTGGTGGT
	1201	GGGGATGCTG	CGGGGATGCT	CGGCGGCTCG	CGGCGGCTCG	CGGCGGCTCG
	1251	TGCGCACCC	CGCGAGGGCA	AGSCTGAGGA	CGCGCGCACT	CAGGTCGGG
	1301	ACGGTCTGCG	GTTTGGGTTT	CATCGGGTGC	TTCCTTTCTT	GGGCGTTTCA
	1351	GCGGGAATCT	CTTTGGCGCA	TGCGCTCTGA		

This encodes a protein having amino acid sequence <SEO ID 212>:

55	1	MMXFXIMLFW	IAGVPVAPFGO	KRLRSXSLGW	LGGSFVSGAS	LWNFLSGVSK
	51	VSGXGKAC	FGSVFRRSGS	RLTGVSTGV	SFLVFSVACP	ASSGCSLSVKA
	101	VSGGGLTRX	FXGAGDGGSP	LPLSSVFSFG	AGDEEAXXC	SGWAASCPIT
	151	FGSGGGLTR	FXGAGCGSVW	RLFSFPGXV	AMTEIANAM	AVLQMSHAR
	201	FRSLGSLIV	FXGAGLILV	GLGSLGSLV	GGTGLGSLV	AVLQMSHAR
60	251	FLYADGGADF	LGNLRLEFGG	DEAHNVYVFG	VGNDFDGTIC	GGAQAQQRSL
	301	DGFCVPSVAF	DVAGSARGSG	DGNVYVHAFG	FLPGTCLNTD	ELFLAFAGBDA
	351	SEQQQVAVVA	DNADDLRVXF	GLVLVLQATG	GGGGDTORRY	VVVXGRRGGS
	401	AVDGGFRADR	RAGDADCADAA	AGEGAGDEGGS	QAGDGVRFYF	HRVFLFAGGS

ORF34a and ORF34-1 show 91.3% identity in 459 aa overlap:

		10	20	30	40	50	60	
	orf34a.pep	MMXFPXIMLFWIAGVPAVFGQKRLSRXSLWGLGGXFFGVSGLVVFSGLVGSXSLGVSGXCAC						
5	orf34-1	MMMPFIMLFWIAGVPAVFGQKRLSRISLWGLGGVFFGVSGLVVFSGLVGSLSL-----GCAC						
		10	20	30	40	50		
	orf34a.pep	FSGVSFRSGSGRTFVGSTGVSLSVFSACAPASSGCLSVXAVSAGCGLTRXFXGAAGDGSP						
10	orf34-1	FSGVSFRSGSGRTFVGSTGVSLSVFSACVPASSGCLSVXAVSAGCGLTRFFLGAAGDGSP						
		60	70	80	90	100	110	
	orf34a.pep	LPLSSVPSGCAGADEEAXXCSGWAASCPPTTFFGQNSVSRGLSVCCGSVVRVLSPPFGXNV						
15	orf34-1	LPLSSVPSGCAGSDEAAWMCWGWAASCPPTTFFGQNSVSRGLSVCCGSAXRVLSPPFLNV						
		120	130	140	150	160	170	
	orf34a.pep	LTMPITANAPMAVIQMSNTARIRSLGVSLKGLPFFAILIVLLGCRAMPSEGGSDGIAESA						
20	orf34-1	LTMPITANAPMAAIQMSNTARIRSLGVSLKGLPGFFAILIVLLGCRAMPSEGGSDGIAESA						
		180	190	200	210	220	230	
	orf34a.pep	LDVVVEGGDDFLYADGGADFLGNLRLFFGGEDAHNVGVAVGNDFDARLCGGADAQQRGA						
25	orf34-1	LDVVVEGGDDFLYADGGADFLGNLRLFFGGEDAHNVGVAVGNDFDARLCGGADAQQRGA						
		240	250	260	270	280	290	
	orf34a.pep	DFGCVPSVAGDVAGSARQGGDGNVXVHAFGLFGTCNLTDDELFLAFGGDLSEQQQVAVVA						
30	orf34-1	DFGCVPSVAGDVAGSARQGGDGNIVVHAFGLFGTCNLTDDELFFAFGGDLSEQQQVAVVA						
		300	310	320	330	340	350	
	orf34a.pep	DNGDLGRVXFGLVVLQAIGAGGGFDTQRHYVVVGVXKAGGSADVDDGFRADRRAADDCADAA						
35	orf34-1	DDGDLGRVAFGLVVLQAIGTGGFDTQRHNVVGLRAGGSADVDDGFRADGASDYCADAA						
		360	370	380	390	400	410	
	orf34a.pep	AEGKAEDGGSQSGADGVRFGRFHRVLPFLGVSDGIALRHAVX						
40	orf34-1	AKGKAENGNGQAGDGVRFGRFHRVLPFLGVSDGIALRHAVX						
		420	430	440	450			
	orf34a.pep	AGDVILLPLSSVPSGCAGSDEAAWMCWGWAASCPPTTFFGQNSVSRGLSVCCGSAXRVLS						
45	orf34-1	AGDGSPLPLSSVPSGCAGSDEAAWMCWGWAASCPPTTFFGQNSVSRGLSVCCGSVVRVLS						
	orf34a.pep	S						
50	orf34ng	PFGLNVLTMPITANAPMAVIQMSNTARIRSLGVSLKGLPFFAILIVLLGCRAMPSEGGSD						
	orf34a.pep	QKSLSRISLWGLGGVFFGVSGLVVFSGLVGSXSE-----CAC						35
55	orf34ng	MMMPFIMLFWIAGVPAVFGQKRLSRISLWGLAGVFFGVSGLVVFSGLVGSLSLGCAC						60
	orf34a.pep	FSGVSFRSGSGRTFVGSTGVSLSVFSACVXGVVRLPVGLSCV-----GRLXLITRFLGA						90
	orf34ng	FSGVSFRSGGGAFFVGSTGVSLSVFSACVP-----VPVNESAAARAASEGR--GLTRFLGA						114
60	orf34a.pep	AGDVILLPLSSVPSGCAGSDEAAWMCWGWAASCPPTTFFGQNSVSRGLSVCCGSAXRVLS						150
	orf34ng	AGDGSPLPLSSVPSGCAGSDEAAWMCWGWAASCPPTTFFGQNSVSRGLSVCCGSVVRVLS						174
65	orf34a.pep	S						175
	orf34ng	PFGLNVLTMPITANAPMAVIQMSNTARIRSLGVSLKGLPFFAILIVLLGCRAMPSEGGSD						234

Homology with a predicted ORF from *N.gonorrhoeae*

ORF34 shows 77.6% identity over a 161aa overlap with a predicted ORF (ORF34.ng) from *N. gonorrhoeae*:

The complete length ORF34ng nucleotide sequence <SEQ ID 213> is:

```

1   ATGATGATGC   CGTTCATAAT   GCTTCCTTGG   ATTGCGGGTG   TGCCTGCCGT
51  GCCGGGTCAA   AAGAGGTTGT   CGAGAATCTC   TTATATGGGT   TTGSCGGGCG
101 TGTTTTTCGG   GGTGTCOCGT   TTGGATATGT   TTTCTTTGGG   GCTTCTCTTT
151 TCCTTTGGGG   TTTCTTTGGG   CTGCGCTGTG   TTTTCCGGTG   TTTCTTTTGG
201 GGGTTCGGGA   TGGGGGGCGT   GTGCGGTTAA   ACGAATCGCG   TGCCCGGGCC
251 TGTTTTCAGC   TGTGTTTCCG   GTGCGGTTAA   TTTCTGGGTG   CGGACGGGGA
301 GCATCCGAGG   GAGCGGTTGT   GACCGGTTGT   TTCTGCGTGT   GTCCGGCTGT
351 ATGAGGTCGG   CTGCGCTTTT   CTCTCTGCGC   GTCCGGCTGT   TCGAGAGGCG
401 CCGTTTGCCA   GCCAGAATTC   GGTTCGCGGG   GGGCTCTCGG   TGTGTTGCGG
451 TTGCGTTTGG   AGGGTTTGTG   CGCGGTTGCG   TTGTAATGTG   CTGACGATGC
501 CTACTGCCAA   TCGCCGCGAT   GCGGTGATAC   AGATGAGCAA   TACGGGCGGT
551 ATCAGGAGTT   TGGGGGTGAG   CCGTAAGGGT   TTGTTGCGTT   TTTTGGCAT
601 TTTGATTGTG   CTTTGGGGGT   GTGCGGCAAT   GCGCTCTGAA   GCGCGTTCAG
651 ACGGCATTGC   CGAGTCAGCG   TTGGACGTAG   TTTTGGTAGA   GGGTAATGAC
701 TTTTGTACG   CCGAGcgtTG   TGCTGACTTT   TTGGGTAACT   TGCCGCTGTT
751 CTTGCGGGGT   GAGGATGCCG   ATACGCTAGG   TTACATTGCC   GTAGGTAATG
801 ATTTTGACGC   GCGCCTGTGT   AGCGGGGCGT   ATGCCACGCA   GcgtgGCGCG
851 GACTTTGGAC   GTGTTCCAAg   TGTGCGCGGC   GATGTGCGCC   GcAGTGGCGG
901 GCAGGGAGGC   GACGGTAAATG   TAGTTGTATA   CGCCTTCGCG   GCGCTTTCG
951 GAACTGTCGA   TCTGACCGAC   GAACTGTTTT   TCSCCTTCGG   TGCGGACTG
1001 TCGAGCAGC   AGCAGGTGGC   GGTGTCAGTA   GACGACGGAG   ATTTGGGCGG
1051 TGTAGCTCTT   GCTTGTGCTG   TTTTGGCGCA   GGTAGGAACG   GCGGTGTGTT
1101 TCGATACCGA   ACGCATTAAC   GTtgtCATCG   GTTtgcgcgc   CGCGGTGTGc
1151 CGGTCGATG   ACGGATTTTG   CGCCGACGGC   GCGCCGCGCG   ACGACTGCGC
1201 TGAAGCAGCC   GCGCAGGGCA   AGGCTGAGGA   GCGCGCAAT   CAGGGTGCAG
1251 ACGGTGTGTG   GTTTGGGTTT   CATCGGGGAC   TTCCTTTCTT   GGGCGTTTCA
1301 GACGGCATTG   CTTTGGCGCA   TGCCGTCGTA
1351

```

This encodes a protein having amino acid sequence <SEQ ID 214>:

```

1   M M M P F I M L F W   I A G V P A V P G Q   K R L S R I S L W G   L A G V F F G V S G   L V W F S L G V S F
51  S L G V S L G C A C   F S G V S F R G S G   W G A F V G S T G V   S L S V F S A C V P   V P V N E S A A R A
101 A S E G R G L T R F   F L G A A G D G S P   L P L S S V P S G C   A G S D E A A N W C   S G W A S C P T A
151 P F G S Q N S V S R   G L S V C G S V W   R V L S P F G L N V   L T M P T A N A P M   A V I Q M S N T A R
201 I R S L G V S L K G   L F G F F A I L I V   L L G C R A M P S E   G S D G I A E S A   L D V L V E G N D
251 F L Y A D G G A D F   L G N L R L F F G G   E D A H N V Y G I A   V G N D F D A R L C   S G A D A Q R G A
301 D F G R V P S V A G   D V A R S A R Q G G   D G N V V Y A P G   G L P G T C N L T D   E L F F A F G D L
351 S E Q Q Q V A V V A   D D G D L G R V A F   G L V I L A Q V G T   G G G E P T Q R H N   V V I G L R A G S
401 A V D D G F C A D G   G P A D D C A E A A   A E G K A E D G N   Q S A D G W F E G   H R G L P F L G V S
451 D G I A L S H A V *

```

ORF34ng and ORF34-1 show 90.0% identity in 459 aa overlap:

		10	20	30	40	50	
orf34-1.pep	M M M P F I M L F W I A G V P A V P G Q N R L S R I S L W L G G V F F G V S G L V W F S L G V S F						-----LGCAC
45 orf34ng	M M M P F I M L F W I A G V P A V P G Q K R L S R I S L W L G G V F F G V S G L V W F S L G V S F S L G V S L G C A C						
		10	20	30	40	50	60
		60	70	80	90	100	110
orf34-1.pep	F S G V S F R G S G R G T F V G S T G V S L S V F S A C V P A S S G C L S V X V A S A G C G L T R F F L G A A G D G S P						
50 orf34ng	F S G V S F R G S G W A F V G S T G V S L S V F S A C V P V P N E S A A R A S E G R G L T R F F L G A A G D G S P						
		70	80	90	100	110	120
		120	130	140	150	160	170
orf34-1.pep	L P L S S V P S G C A G S D E A A M W C S G W A A S C P T P F G S Q N S V S R G L S V C G S A X R V L S F F G L N V						
55 orf34ng	L P L S S V P S G C A G S D E A A M W C S G W A A S C P T A P F G S Q N S V S R G L S V C G S V W R V L S F F G L N V						
		130	140	150	160	170	180
		180	190	200	210	220	230
orf34-1.pep	L T M P I A N A P M A A I Q M S N T A R I R S L G V S L K G L F G F F A I L I V L L G C R A M P S E G G S D G I A E S A						
60 orf34ng	L T M P T A N A P M A V I Q M S N T A R I R S L G V S L K G L F G F F A I L I V L L G C R A M P S E G G S D G I A E S A						
		190	200	210	220	230	240

		240	250	260	270	280	290
5	orf34-1.pep	LDVVVLVEGDDFLYADGGADFLGNLRLFFGGEDAHNVGYVAVGNDFDARLCGGADAQQRGA					
	orf34ng	LDVVVLVEGDDFLYADGGADFLGNLRLFFGGEDAHNVGYVAVGNDFDARLCGGADAQQRGA					
		250	260	270	280	290	300
10	orf34-1.pep	DFGCVFVSAGDVAGSARQGGDGNVIVVHAFGGLFCTCNLTDELFFAFGGDLSEQQQVAVVA					
	orf34ng	DFGRVFSVAGDVARSARQGGDGNVIVVYAFGGGLFCTCNLTDELFFAFGGDLSEQQQVAVVA					
		310	320	330	340	350	360
15	orf34-1.pep	DDGDLGRVAFGLVVLVAQIGTGGGFDTRHNVVGLRAGGSADVDDGFADGGASDYCADAA					
	orf34ng	DDGDLGRVAFGLVVLVAQVGTGGGFDTRHNVVGLRAGGSADVDDGFADGGASDYCADAA					
		370	380	390	400	410	420
20	orf34-1.pep	AKGKAENGNGQAGDGVRFGRHVLFFLGVSDGIALRHAVX					
	orf34ng	AEGKAEDGNGQAGDGVRFGRHVLFFLGVSDGIALRHAVX					
		430	440	450	460		

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 26

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 215>:

```

1  ATGAAAACCT TCTTCAAAAC CCTTCCGCC GCGGCACTCG CGCTCATCCT
51  CGCCGCGCTGC GGATT CAAA AAGACAGCGC GCGCGCCGCA TCCGCTTCTG
101 CGCGCGCGCA CAACGGCGCG GGTAAAGAA GAATGCTCT TCGCGACGAC
151 CGTCGGCGACT TCGCGCGATA TGGTCAAGA ACAATCCAA GCGAGCTGG
201 AGAAAAAAGG CTACACCGTC AACTGGTGC AGTTTACCGA CTATGTACGC
251 CCGAATCTGG CATTGCTGA GCGGAGTTG

```

This corresponds to the amino acid sequence <SEQ ID 216; ORF4>:

```

1  MKTFFKTLA AALALILAA G.QKDSAFAP SASAADNGA AKKEIVFGTT
51  VGDFGDMVKE QIQAELEKKG YTVKLVEFTD YVRPNLALAE GEL

```

Further sequence analysis revealed the complete nucleotide sequence <SEQ ID 217>:

```

1  ATGAAAACCT TCTTCAAAAC CCTTCCGCC GCGGCACTCG CGCTCATCCT
51  CGCGCGCTGC GGCGGTCAA AAGACAGCGC GCGCGCCGCA TCCGCTTCTG
101 CGCGCGCGCA CAACGGCGCG GCGAAAAAG AAATCGTCTT CGGCAACGAC
151 GTGCGGCACT TCGCGGATAT GGTCAAAGAA CAATTCGCA GCGAGCTGGA
201 GAAAAAGGC TACACCGTCA AACTGGTCSA GTTTACCGAC TATGTACGCC
251 CGAATCTGGC ATTGCTGAG GCGAGTTGG ACATCAAGT CTTCACACAC
301 AAACCTATC TTGACGACTT CAAAAAGAA CACATCTGG ACATCACCGA
351 AGTCTTCCAA GTGCGGACGC CGCTTTTGGC ACTGTACCGC GGCAGCTGA
401 AATCGGTGGA AGAAGTCAA GCGGCGACGA CCGTATCCGC GCCCAAGAC
451 CGGTCCAACT TCGCGCGCGT CTTGGTGATG CTGACGAAC TGGGTTGGAT
501 CAAACTCAAA GACGCGCATCA ATCCGTTGAC CGCATCCAAA GCGGCACTCG
551 CCGAGAACCT GAAAAACATC AAAATCGTGC AGCTTGAAAG CGCGCAACTG
601 CCGCGTAGCC GCGCGGACGT GGATTTTGCC TCGCTCAAGC GCAACTACGC
651 CATAGCAGC GGCATGAAGC TGACCGAAGC CTTGTTCCAA GAACCGAGCT
701 TTGCTATGT CAACTGGTCT CGCGTCAAAA CGCGCGACAA AGACAGCCAA
751 TGGCTTAAAG ACGTAACCGA GGCCTATAAC TCCGACGCGT TCBAAGCCTA

```

801 CGCGCACAAA CGCTTCGAGG GCTACAAATC CCCTGCCGCA TGGATGAAG
851 GCGCAGCCAA ATAA

This corresponds to the amino acid sequence <SEQ ID 218; ORF4-1>:

5 1 MKTFFKTL~~SA~~ AALALILAA~~C~~ GGQKDSAPAA SASAADNGA AKKEIVFGTT
51 VGDFGDMVKE QIQAELEKKG YTVKLVEFTD YVRNPLALAE GELDINVFOH
101 KFYLDDEFKE HNLDTITVFO VPTAPLGLYP GKLSLEEVK DGSTVSPFND
151 PSNFARVLVM LDELGXIKLK DGINPLTASK ADIAENLNKI KIVLELAQAL
201 PRSRADVDFA VVNGNYAISS GMKLTEALFQ EPSFAYVNWS AVKTADKDSQ
251 WLKDVTETAYN SDAFKAYAHK RFEYKSPAA WNEGAAK*

10 Computer analysis of the amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF4 shows 93.5% identity over a 93aa overlap with an ORF (ORF4a) from strain A of *N. meningitidis*:

		10	20	30	40	50	59
15	orf4.pep	<u>MKTFFKTL</u> SA AR LA L ILAA C GQKDSAPAA S ASAAADNGAAKKEIVFGTTVGDFGDMVKE					
	orf4a	<u>MKTFFKTL</u> SA AR LA L ILAA C GQKDSAPAA S ASAAADNGAAKKEIVFGTTVGDFGDMVKE					
		10	20	30	40	50	60
20	orf4.pep	60	70	80	90		
		QIQAELEKKG YTVKLVEFTDYVRNPLALAE GEL					
	orf4a	XIQPELEKKG YTVKLVEFTDYVRNPLALAE GELINVQHXHYLDKXKHNLDTITVXQ					
		70	80	90	100	110	120
25	orf4a	VPTAPLGLYPGKLSLXXVKGSTVSPAPNDPXXFXRVLVMDLDELGXIKLKDXIXXXXXXX					
		130	140	150	160	170	180

The complete length ORF4a nucleotide sequence <SEQ ID 219> is:

30 1 ATGAAACCT TCTTCAAAAC CCTTCGCGC GCGGCACTCG CGCTCATCCT
51 GCGCGCTCTG GCGCGTCAAA AAGATGACG GCGCGCCGCA TCGCGTCTCG
101 CCGCGCCGCA CAACGCGCGG GCGAATAAAG AATCGTCTT CGCGCAGACC
151 CTGCGGCACT TCGCGGATAT GGTCAAAGAA CANATCCAAC CCGAGCTCGA
201 GAAAAAAGGC TACACCGTCA AACTGGTCTGA GTNATCCGAC TATGTCCGCN
251 CGAATCTGGC ATTGGCTGAG GCGGAGTTGG ACATCAACGT CTNCAACAC
35 301 ANACNCTATC TTGACGACTN CAAAAANAA CACAACTCGG ACATCACCN
351 AGTCTTNCAT GTGCGGACCG CGCCTTTGGG ACTGTACCGC GGCAAGCTGA
401 AATCGTCTGA NNAAGTCAAA GANGGACGCA CCGTATCCGC GCCCAACGAC
451 CCGTNNNACT TCGNCCGCGT CTGGTGATG CTCGACGAAC TGGGTNGAT
501 CAAACTCAAA GACNGCATCA NNNNGNNGNN NNNANCNANA NNGANANNN
40 551 NNNNNNNNT NNNNNNNNN NNNNNCNCG NNNNNNNNN NNNNNNNNN
601 NCGNNNNNN NNGCNNNNT NNNNNNNNN NNNNNNNNN NNNNNNNNN
651 NANNANNAGC GGCATGAAGC TGACGGAAGC CCTGTTCCAA AGACCGAGCT
701 TTGCTTATGT CAACCTGCTCT GCGCTCAAAA CCGCGGACAA ACACAGCCTA
45 751 TGGCTTAAGG ACGTAACCGA GGCCTATAAC TCCGAGCGGT TCACAGCCTA
801 GCGGCACAAA CGCTTCGAGG GCTACAAATC CCCTGCCGCA TGGATGAAG
851 GCGCAGCCAA ATAA

This is predicted to encode a protein having amino acid sequence <SEQ ID 220>:

50 1 MKTFFKTL~~SA~~ AALALILAA~~C~~ GGQKDSAPAA SASAADNGA AXKEIVFGTT
51 VGDFGDMVKE XIQPELEKKG YTVKLVEFTD YVRNPLALAE GELDINVXQH
101 XXYLDLXKXK HNLDTITVXQ VPTAPLGLYP GKLSLXVKV XGSTVSPAND
151 PXFXRVLVM LDELGXIKLK DXIXXXXXXX XXXXXXXXXX XXXXXXXXXX
201 XXXXAXXXXX XXXXXXXXXX GMKLTEALFQ EPSFAYVNWS AVKTADKDSQ
251 WLKDVTETAYN SDAFKAYAHK RFEYKSPAA WNEGAAK*

A leader peptide is underlined.

Further analysis of these strain A sequences revealed the complete DNA sequence <SEQ ID 221>:

```

1   ATGAAAACCT TCTTCAAAC CTTTCCGCC GCGCAGCTCG GGCTCATCCT
51  CGCGCGCTGC GCGCGTCARA AAGATAGCGC GCGCGCCGCA TCCGCTTCTG
101 CGCGCGCCGA CAACGGCGCG GCGAAAAAAG AATCGTCTT CGGCAGCGAC
151 GTCCGCGGACT TCGGGGATAT GGTCAAAGAA CAATCCACAC CCGAGCTGGA
201 GAAAAAAGGC TACAACGTC AACTGGTCGA GTTACCGSAC TATGTGCGCC
251 CGAATCTGCG ATTGGCTGAG GCGAGTTGG ACATCAACGT CTTCACACAC
301 AAACCCCTATC TTGACGACTT CAARAAGAA CCAATCTGG ACATCACCGA
351 AGTCTTCCAA GTCCGCGCG GCGCTTTGGG ACTGTACCGG GGCACGCTGA
401 AATCTCTGGA AGAATCAAA GACGCGACGA CGTATCCGCG GCCCAACGAC
451 CCGTCCCACT TCGCCCGCGT CTTGGTGATG CTCGACGAAC TGGGTTCGAT
501 CAACTCAAA GACGCGATCA ATCCGCTGAC CGCATCCAAA GCGGACATTG
551 CCGAAAACCT GAAAACATC AAAATCGTCG AGCTTGAGCG CGCGCACTG
601 CGCGGTAGCC GCGCGGACGT GGATTTTGCC GTCGTCAACG GCAACTACGC
15 651 CATAAGCAGC GGCATGAAGC TGACCGAAGC CCTGTTCCAA GAACCGAGCT
701 TTGCGTATGT CAACTGGTCT GCGCTCAAAA CCGCGACAAA AGACAGCCAA
751 TGGCTTAAGC ACGTAACCGA GGCCTATAAC TCCGACGCGT TCAAAGCCTA
801 CGCGCACAAA CGCTTCGAGG GCTACAAATC CCTCGCCGCA TGGAAATGAAG
851 GCGAGCGCAA ATAA

```

This encodes a protein having amino acid sequence <SEQ ID 222; ORF4a-1>:

```

1   MKTFFKTLSA AALALILAAC GGQKDSAPAA SASAAADNGA AKKEIVFGTT
51  VGDFGDMVKE QIQPELEKKG YTVKLVEFTD YVRPNLALAE GELDINVFQH
101 KPYLDDEFKE HNLDTIEVFQ VFTAPLGLYP GKIKSLEEVK DGSVTSAPND
151 PSNFARVLVM LDELGWIKLK DGINPLTASK ADIAENLNKI KIVLEEAQNL
201 PRSRADVDFV VVNGNYAIS SSQMKLALFQ EPSFAYVNW AVKTADKDSQ
251 WLKDVTEAYN SDAFKAYAHK RFEQYKSPAA WNEGAAK*

```

ORF4a-1 and ORF4-1 show 99.7% identity in 287 aa overlap:

		10	20	30	40	50	60
30	orf4a-1	MKTFFKTLSA	AALALILAAC	GGQKDSAPAA	SASAAADNGA	AKKEIVFGTT	VGDFGDMVKE
	orf4-1	MKTFFKTLSA	AALALILAAC	GGQKDSAPAA	SASAAADNGA	AKKEIVFGTT	VGDFGDMVKE
		10	20	30	40	50	60
35	orf4a-1		70	80	90	100	110
	orf4-1		70	80	90	100	110
			70	80	90	100	110
40	orf4a-1		130	140	150	160	170
	orf4-1		130	140	150	160	170
			130	140	150	160	170
45	orf4a-1		190	200	210	220	230
	orf4-1		190	200	210	220	230
			190	200	210	220	230
50	orf4a-1		250	260	270	280	
	orf4-1		250	260	270	280	
			250	260	270	280	

Homology with an outer membrane protein of *Pasteurella haemolytica* (accession q08869).

ORF4 and this outer membrane protein show 33% aa identity in 91aa overlap:

[illegible]

Homology with a predicted ORF from *N.gonorrhoeae*

ORF4 shows 93.6% identity over a 94aa overlap with a predicted ORF (ORF4.ng) from *N.*

20	<i>gonorrhoeae:</i>							
	orf4nm.pep	MKTFFKTL	SAAALALI	LILAACGQXKD	SAPAPA	10	20	30
		: :	: :	: :	: :			
25	orf4ng	RANAVXTNPNDGRTPCLSFLETATTSGENMKTFKKTSTASLALIILAACGGQKDSAPAPA	200	210	220	230	240	250
	orf4nm.pep	SASA-AADNGAAKKEIVFGTTVGDFGMVKEIQIQAILEKKGYTVKLVEFTDYVRPNRLALA	40	50	60	70	80	89
		: : :	:~ ~ ~ ~ ~	Q I A L E K K G Y T V K L V E F T D Y V R P N R L A L A				
30	orf4ng	AAAAPSADNGAAKKEIVFGTTVGDFGMVKEIQIQAILEKKGYTVKLVEFTDYVRPNRLALA	260	270	280	290	300	310
	orf4nm.pep	EGL	EGEL	I				
35	orf4ng	EGELDINVFQHCKPYLDDEKKEHNLDITEAFQVPAPTPLGPLPGPKLSLEEVDKGdstvsapn	320	330	340	350	360	370

The complete length ORF4ng nucleotide sequence <SEQ ID 223> was predicted to encode a protein having amino acid sequence <SEQ ID 224>:

40	1	KTFFFKTLST	ASIALILPAC	GGKDSAPAA	SAAAPSADNG	AAKKEIVFGT
	51	TGVDGDMVY	QIQAELEKK	TVTKLIVFE	DVVRPNLALA	EGELDNVFO
	101	HKPYLDFDKK	HNHLDTGEAF	QVPTAPLPLF	PGKIKLLEEV	KDGTSTVSAFN
	151	DPSNFARALV	LMNLELQWIKL	KDGINNGLTS	GADKAENIKN	IKIVLELAQA
45	201	LPRSRADVDF	AVSNGVNYAIS	SGMKLPEALF	QEPSFAYVNW	SAVKTADKDS
	251	QLKQDVTEAY	NNDNFKAYAH	KRFEGYKYVA	AWNNGEAAK*	

Further analysis revealed the complete length ORF4ng DNA sequence <SEQ ID 225> to be:

	1	atgAAACCTC	TCTTCAAACG	actcttcgcg	gccgcacCTGG	CGCTCATCTCT
	51	ccAGCGCTCG	ggCGctcaAA	AGAGCAAGCGC	CGCGcgagcc	ctctgcCGCGC
	101	CCCTCTCTCG	CGATAAACGcg	cgCGGCAAAA	AGAAATcgt	ctTCGGCAGC
50	151	Acctctggcg	actcttcggcG	TALggTCAAA	GAGCAAAATCG	AagcCGAgct
	201	gGCAAAAAA	GcctACACCG	tCAAttgtg	cgaatttacc	gactatgtGC
	251	gcCGGATATC	GGATCTGGCG	TAGGAGCGAT	TGGACATCAA	CGTCTTCBA
	301	CAACAAACCT	ATCTTGAACG	TTTCAAAAAA	GAACAAACCC	TGGACATCAC
55	351	CGAAGCGCTT	CAAATGCGCA	CGCCCGCTTT	GGGACTGTAT	CGGGCAAAAC
	401	TGAATATCTC	GGAAGAAGTC	AAAGACGGCA	CACCGCTATC	CGCGCCCAac
	451	gAcctcgTGA	ACTTTCGACG	CGCTTGGTGG	ATGCTGAAGG	AACTGGGTGG
	501	TATCAAACTC	AAAGACGGCA	TCAATTCGCT	GACCGCATCC	AAAGCGGCAC
	551	GTGCGAGAAA	CTCGAAAAAC	TGTCAAAATG	TGCAGATTGA	AGCGCGCAAC
	601	CTGCGCAGCA	AGCGCGCGCA	CGTGGAATTT	CGCGTCGTCA	AGCGCAACTA
60	651	CGCCATAAGC	GACGGCATCA	AGCTGACCGA	AGCCCTGTTT	CAAGAGCGCA

701 GCTTTGCCTA TGTCACCTGG TCTGCCgtcA AAACCGCCGA CAAAGACAGC
 751 CAATGGCTTA AAGACGTAC CGAGGCCTAT AACTCCGACG CGTTCAAAGC
 801 CTACCGGCAC AAACGCTTCG AGGGCTACAA ATACCTCGCC GCATGGAATG
 851 AAGGGCGAGC CAAATAA

5 This encodes a protein having amino acid sequence <SEQ ID 226; ORF4ng-1>:

1 MKTFFKTLISA AALALILAAC GGQKDSAPAA SAAAPSADNG AAKKEIVFGT
 51 TVGDFGDMVK EQIQAELEKK GYTVKLVEFT DYVRPNLALA EGELDINVFO
 101 HKPYLDFFKK EHNLDITEAF QVTPAPLGLY PGKLSLEEV KDGSTVSAPN
 151 DPSNFARALV MLNELGWIKL KDGINPLTAS KADIAENLKN IKIVELEAAQ
 201 LPRSRADVDF AVVNGNYAIS SGMKLTEALF QEPSPFAYVNW SAVKTADKDS
 251 QWLKDVTEAY NSDAFKAYAH KRFEQYKYPAAWNEGAAK*

This shows 97.6% identity in 288 aa overlap with ORF4-1:

		10	20	30	40	50	59
15	orf4-1.pep	MKTFFKTLISAAALALILAACGGQKDSAPAAASASA-AADNGAAKKEIVFGTTVGDFGDMVK					
	orf4ng-1	MKTFFKTLISAAALALILAACGGQKDSAPAAASAAAPSADNGAAKKEIVFGTTVGDFGDMVK					
		10	20	30	40	50	60
20	orf4-1.pep	60	70	80	90	100	110
	orf4ng-1	EQIQAELEKKGYTVKLVEFTDYVRPNLALAEGLDINVFOHKPYLDFFKKEHNLDITEAF					
		70	80	90	100	110	120
25	orf4-1.pep	120	130	140	150	160	170
	orf4ng-1	QVTPAPLGLYPGKLSLEEVKDGSTVSAPNDPSNFARVLMLDELGWIKLKDGINPLTAS					
		130	140	150	160	170	180
30	orf4-1.pep	180	190	200	210	220	230
	orf4ng-1	KADIAENLKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSPFAYVNW					
		190	200	210	220	230	240
35	orf4-1.pep	240	250	260	270	280	
	orf4ng-1	SAVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEQYKYPAAWNEGAAKX					
		250	260	270	280		
40	orf4-1.pep	SAVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEQYKYPAAWNEGAAKX					
		250	260	270	280		

In addition, ORF4ng-1 shows significant homology with an outer membrane protein from the database:

45	ID	LIP2_PASHA	STANDARD;	PRT;	276 AA.
	AC	Q08869;			
	DT	01-NOV-1995	(REL. 32, CREATED)		
	DT	01-NOV-1995	(REL. 32, LAST SEQUENCE UPDATE)		
	DT	01-NOV-1995	(REL. 32, LAST ANNOTATION UPDATE)		
50	DE	28.2 KD OUTER MEMBRANE PROTEIN PRECURSOR.			
	SCORES	Init1:	279	Initn:	416
		Opt:	494		
		Smith-Waterman score:	494;	36.0% identity in	275 aa overlap
55	orf4ng-1.pep	MKTFFKTLISAAAL--ALILAACGGQKDSAPAAASADNGAAKKEIVFGTTVGDFGDM			
	lip2_pasha	MNFKKLLGVALVSALALTACKDEKAQAPATTA---KTENKAPLK---VGVMTGPFAQIM			
		10	20	30	40
60	orf4ng-1.pep	60	70	80	90
		VKEQIQAELEKKGYTVKLVEFTDYVRPNLALAEGLDINVFOHKPYLDFFKKEHNLDITE			
		100	110		

	lip2_pasha	TEVAVKIAKEKYGLDVELVQFTEYTPQPNAAHLSKDLDDANAFQTVFPYLEQEVKDRGYKLAI	60	70	80	90	100	110
5	orf4ng-1.pep	120 130 140 150 160 170 AFQVPTAPLGLSLPGKLSLEEVVDKGSVTSAPNDPSNFARALVMLNELGWMILKLDGINPLT :: : : : : : : : : ::: : :						
	lip2_pasha	120 130 140 150 160 170 IGNTLVWPIAAYSKKIKNISLKDGAIVAI PNNASNTARALLLQAHGLLKLKDPKN-VF						
10	orf4ng-1.pep	180 190 200 210 220 230 ASKADIAENLNKIKIVLEAAQLPRSRADVDFAVVNGNYAISSGMKLTE--ALFQEPSPFA :: : :: :: :: ::: : : : :						
	lip2_pasha	180 190 200 210 220 230 ATENDI IENPKNIKIVQADTSLLRMLDDVELAVINNTYAGQAGLSPDKDGIIVSEKSDSP						
15	orf4ng-1.pep	240 250 260 270 280 289 YVNWSAVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEQYKYPAWNEGAARKX : : : : : ::::: :						
	lip2_pasha	240 250 260 270 YVNLVVSREDKDDPRLQTTFVKSFTTEVFQEALKLFNGGVVKGW						
20								

Based on this analysis, including the homology with the outer membrane protein of *Pasteurella haemolítica*, and on the presence of a putative prokaryotic membrane lipoprotein lipid attachment site in the gonococcal protein, it was predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF4-1 (30kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figures 8A and 8B show, respectively, the results of affinity purification of the His-fusion and GST-fusion proteins. Purified His-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result), Western blot (Figure 8C), FACS analysis (Figure 8D), and a bactericidal assay (Figure 8E). These experiments confirm that ORF4-1 is a surface-exposed protein, and that it is a useful immunogen.

Figure 8F shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF4-1.

Example 27

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 227>:

```

1   CCTCGTOSTC   CTCGGCATGC   FCCAGTTTCA   AGGGGCGGATT   TACTCCAAGG
51  CGGTGGAAACG   TATGCTCGGC   ACGGTCATCG   GGCTGGGCGCG   GGGTTTGGCG
101 GTTTTATGGC   TGAACCAAGCA   TTATTTCAC   GGCAACCTCC   TCTTCTACCT
151 CACCGTCGGC   ACGGCAAGCG   CACTGGCCGG   CTGGGCGGGCG   GTCGGCAAAA
201 ACGGCTACGT   CCTmTgCTG   GCAGGGCTGA   CGATGTGTAT   GCTCATCGGC
251 GACAAACGCA   GCGAATGGCT   CGACAGCGGA   CTCATGCGCG   CCATGAACGT
301 CCTCATCGGC   GYGCCCATCG   CCATCGCGCG   CGCGAAACTG   CTGCCGCTGA
351 AATCCACACT   GATGTGGCGT   TTTCATGCTG   CGGACAACTG   GGCCGACTGC
401 AGCAAAATGA   TTCCGGAANT   CAGCAACGGC   AGGCGCATGA   CCGCGAAGC
451 CCTCGAGSAG   AACATGGCGA   AATCAACGCA   CGCATGGTCA   CGCATGGTCA
501 AAGCCGCGAG   CCATCTCGGC   GCCACATCGG   GCGAAGCTG   CATCAGCCCC

```

551 GGCATGATGG AAGCCATGCA GCACGCCAC CGTAAATATG TAACACCAAC
601 CGAGCTGCTC CTGACACGCG CCACTCTCCC AACTCAACG
651 GCAGCGAATAT CCGGCTGCTT GACGCCACTT TCACATGCTT CCACAC....
701 GC AGACAGGCCG CGGCATCGC
751 CATCGACACC GCATCAATCC CGGATATGCA AGCCCTCGCC GAACACATCC
801 ACTACCAATG CAGCGGCTTC TCTGGCTGTA CACGAGTAT CGCTCAGGAA
851 ATTTCGCGCC TCGTCACTCT GTCTCAACGC ACCCGCCGCA AATGGCTGGA
901 TGCCACAGAA GCGCCACAAAG TGCGCCAAAG CCTGCTTGA

This corresponds to the amino acid sequence <SEQ ID 228; ORF8>:

10	1PRRP	RHAFVSRGDL	LQGGTYARH	GHRAGRGFGR	FMAEPALFPR
	51	QPPLLPHRRH	GKRTGRFGG	RQKRLRFXAG	RADDVYAIHR	QQRMARQRT
	101	HAEHRPHRR	GKRRHRRGTA	AAETHDVAF	HACRQPGRIKQ	QNDCRNQORQ
	151	ADHPTPRGE	HGEANPNQRT	HGQKPGPAXR	HIGKRLHQPR	HGDGSHAARPJG
	201	XNRGTHRAAP	DHRRQAALISQ	TQQRQNPASR	PLPHTAN..Q
15	251	PRPHPHRRH	HPQRTGSPRR	TPPLPMAGLP	LAQHRYASGN	FRPRHPAATH
	301	PPPMACGPRT	PTPAKPKA*			

Computer analysis of this amino acid sequence gave the following results:

Sequence motifs

20 ORF8 is proline-rich and has a distribution of proline residues consistent with a surface localization. Furthermore the presence of an RGD motif may indicate a possible role in bacterial adhesion events.

Homology with a predicted ORF from *N.gonorrhoeae*

ORF8 shows 86.5% identity over a 312aa overlap with a predicted ORF (ORF8.ng) from *N.*

gonorrhoeae:

25	orf8.ng	1	MDDRRLLRPRHPAPVPRRLDLQGGTYARYVGHGRAGRGFRGMAEPALFPR	50
	orf8.pep	1PRRPRHPAPVSRGDLQGGTYARHGHRAGRGFRGMAEPALFPR	44
30	orf8.ng	51	QPPLLPDHRHGKRTGRLGGGQKKRLRPVVGADDVVAHRRQRQRMARQRP	100
	orf8.pep	45	QPPLLPPIRRHGKRTGRLGGGQKKRLRPXGAGRADVVAHRRQRQRMARQRT	94
	orf8.ng	101	DARDERPHRRRHRCRRQTAAAEIHTDVAFHACRQPGRLQQNDCRNQQRQ	150
35	orf8.pep	95	HARHERPHRRGHRHRRRTQAAAEIHTDVAFHACRQPGRMQQNDCRNQQRQ	144
	orf8.ng	151	AYDARTFGAEYQGONAPNQTHGQKQPQPRRHIGRKPHQPLHDGSHAARPP	200
	orf8.pep	145	AHDPRTPRGEHGENAPNQTHGQKQPQSRRHIGRKLHQPHRHDGSHAARPP	194
40	orf8.ng	201	QNRQHHRRAAPDHRQQAISQTQRQRNPAARPLHTAPNRPATNRRPHQRQ	250
	orf8.pep	195	XNRQHHRRAAPDHRQQAISQTQRQRNPAARPLHTAPN.....P	244
45	orf8.ng	251	TRPPPHRRHRHQPTGSPRTPLPMAGFLAQHQYASGNFRPRHPAATH	300
	orf8.pep	245	TRPPPHRRHRHQPTGSPRTPLPMAGFLAQHRYASGNFRPRHPAATH	294
	orf8.ng	301	PPQAGCPRTPTPAPKPA* 319	
50	orf8.pep	295	PPQAGCPRTPTPAPKPA* 313	

The complete length ORF8ng nucleotide sequence <SEQ ID 229> is predicted to encode a protein having amino acid sequence <SEQ ID 230>:

5
 1 MDRDDLRRP RHAPVPRDL LQRGGTYARY GHRAGRGFR FMAEPALFPR
 51 QPILLPDHRH GKRTGLGGG RQKRLRPVVG GADDVHAHRH QRQRMARQRP
 101 DARDERPHRR RHRHCRROTA AAEIHTDVAF HACRQGRGLQ QNDCRNQORQ
 151 AYDARTFGAE YGQNAPNQRT HGQKPQPPRR HIGRKKHQPL HDGSHAARPP
 201 QNRQHHRAAP DHRRAALISQ TQRQNPFAAR PFLHTAPNRP ATNRRHQORQ
 251 TRFPHHRHR HQRATGSERR TFFLFMAGFF LAQHQYASGN FRFPHHPATH
 301 FFGMGCPRT FTPAPKPA*

Based on the sequence motifs in these proteins, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 28

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 231>:

1 ..GAAATCAGCC TCGGTCGGA CNACAGGCGG GTTTCGGTGN CGAAGCGGG
 51 GGATTTCGAA CGTTTTCTGC TGTGGACGG CGGCACACAG CGGCTCAAGT
 101 GGGCGTGGGT GGAAACGGC ACGTTCGCAA CCGTCGGTAG CGCGCGCTAC
 151 CGCGATTGTG CGCCTTTGGG CGCGAGGTGG CGGAAAAGG CGGATGGAAA
 201 TGTCCGCATC GTCGGTTGGC CTGTGTGGG AGAATTCAA AAGCACACAG
 251 TGCAGGAACA GCTCGCCCGA AAAATCGAGT GGCTGCGGTC TTCCGACAG
 301 GCTTTGGCA TACGCAACCA CTACCGCCGC CCGAAGAAC ACGGTTTCGA
 351 CCGCTGGTTC AACCGCTGG GACCGCGCGC CTTACGCGC AAGGCGTCGG
 401 TCGTCTCAG TCGCGCAAG CGGCTAACGG TTGACGCGCT CACCGATGAC
 451 GGACATATC TCGGAGA GG AACCATCATG CCGGTTTCC ACCTGATAGA
 501 AGAATCGCTC GCGCTCCGAA CCGCACCCT CAACCGGCAC GCGGTAAGC
 551 GTTATCCTTT CCGACCGG..

This corresponds to the amino acid sequence <SEQ ID 232; ORF61>:

1 ..EISLRSDKRP VSVXKRROSE RFLLLDGGNS RLKAWWVNG TFAIVGSAPY
 51 RDLSPGLAEW AEKADGNVRI VGCAVCGEFK KAQVQELAR KIEWLPSSAQ
 101 AXGIRNHYRH PEEHSGDRWF NALGSRFRFR NACVVVSOGT AVTVDALDD
 151 GHYLGXGTIM PGPHLMKESL NVRTANLNHR AGKRYPFPT..

Further work revealed the complete nucleotide sequence <SEQ ID 233>:

1 ATGACGGTTT TGAAGCTTTC GCACTGGCGG GTGTGGCGG AGCTTGCCGA
 51 CGGTTTCCGG CAACACGTCT GCGAAGTGGC GCGTATGGCG GATATGAAGC
 101 CGCAGCMCT CACGCTTTT TGGCAGCAGA TGGCGGCGCA CATACGCGGG
 151 CTGTTCCGCC AACACGACGG CTATTGGCGG CTGTGCGCC CATTTGCGGT
 201 TTTCGATGCC GAAGGTTTGC GCGAGCTGGG GGAAGCTGCG GGTTTTCA
 251 CGGCAITGAA GCACGAGTGC GCGTCCAGCA ACGACGAGAT ACTGGAATTG
 301 GCGCGGATTG CGCCGGACAA GCGCACAAA ACCATATGCG TGACCCACTC
 351 GCAAAGTAAG GGCAGGGGGC GGCAGGGGCG GAAGTGGTGC CACCGTTTGG
 401 CGGAGTGCTC GATGTTCACT TTTGGCTGGG TGTTTGACCG GCGCGAGTAT
 451 GAGTTGGGTT CGCTGTCGCC TGTGTCGGCA GTGGCGTGTG GCGCGCCTT
 501 GTGCGGTTTA GGTTTGGATG TGCAGATTAA GTGGCCCAAT GATTTGGTTG
 551 TCGGACGCGA CAATTTGGGC GGCATTCTGA TTGAACGGT CAGGACGGCG
 601 GGCAAAACGG TTGCGTGTGT CGGTATCGCG ATCAATTTTG TCCTGCCCAA
 651 GGAAGTAGAA AATGCGCGTT CGGTGCAATC GGTGTTTCAG ACGGATCGCG
 701 GCGCGGCGCA TGGCGATGCC CGCGTGCTGC TGGAAACGCT GTTGGTGGAA
 751 CTGACGCGG TCTTTTCGCA ATATGCGCGG GACGATTATG CGCCTTTTGT
 801 GCGGAATAT CAGGCTGCCA ACCGCGACCA CGGTGAAGCG GTTATGCTGT
 851 TGCAGACGCG GAAACCCGTC TTCGAAGCCA CGGTTAAGCG CCGGACGA
 901 CAAGGCGGTT TGCACTTGGG AACGGCAGAG GGCAAAACGA CGGTGCTGAT
 951 CGCGGAAATC AGCCTCGGGT CCGACGACAG CCGGTTTCC TGCAGAAC
 1001 GCGCGGATTC GGAACGTTTT CTGCTGTTGG ACAGGCGCAA CAGCGCGCTC
 1051 AAGTGGCGGT GGGTGGAAAA CGGCACGTTT GCAACCGTGC GTAGCGCGCC
 1101 GTACCGCGAT TTGTGCGCTT TGGGCGCGGA GTGGGCGGAA AAGCGGATG
 1151 GAAATGTCGG CATCGTCGGT TGCCTGTGTG GCGGGAATTT CAAAAAGGCA
 1201 CAAGTCAGAG AACAGCTGCG CCGAAAAATC GAGTGGCTGC CGTCTTCCGC
 1251 ACAGGCTTTG GGCATACGCA ACCACTACCG CCACCCGAA GAACAGCGTT
 1301 CCGACCGCTG GTTCAACGCC TTGGGCGAGC GCGCTTTCAG CCGCAACGCC

1351 TGCGTGTGTG TCAGTTGCGG CACGGCGGTA ACGGTTGACG CGCTCACCGA
 1401 TGACGGACAT TATCTCGGGG GAACCATCAT GCCCGTTTC CACCTGATGA
 1451 AAGAAATGCT GCCTGTCGGA ACCGCCAACC TCACACGCGA CGCCGCTAAG
 1501 CGTATATCTT ACGCGGCAT CCGCTGATAT CGCGCATATAT
 1551 GATGCGGCTT TCCGGCTCGG TTATGATGAT GCACGGGCGT TTGAAGAAGA
 1601 AATCCGGGCG GGGCAAGCCT GTCCATGTCA TCATTACCGG CGCGCGCGCG
 1651 GCAAAATGTC CCGAAGCCCT CGCCCTGCA TTTTGGCGG AAAATACCGT
 1701 GCCTGTGGCG GACAACCTCG TCATTATACG GTTGTGTAAC ATGATTGCGG
 1751 CCGAAGGCAG GGAATATGAA CATATTTAA

10 This corresponds to the amino acid sequence <SEQ ID 234; ORF61-1>:

1 MTVLKLSHRV VLAELADGLP QHVSQALARMA DMKPOQLNGF WQMPFAHNRG
 51 LLRQHDGYWR LVRPLAVFDA EGLRELGERG GFQTKALHEC ASSNDEILEL
 101 ARIAPDKAHK TICVTHLQSK GRGRQGRKWS HRIGCELMFS FGWVDFRPOY
 151 ELGSLSPVAA VACRRALSRL GLDVQIKWPN DLVVGRDKLG GILLETVRTG
 201 GKTAVVVGIG INFVLPKEVE NAASVQSLFP TASRRGNADA AVLETLLEVE
 251 LDVAVLLQYAR DGFAPFVAEY QAANRDHCKA VLLLRDGETV FEGTVKGVGDG
 301 QGVHLHETAE GKQTVVSSEI SLRSDDRPVV VPKRRDSERF LLLDGGNSRL
 351 KWAUVVNGTF ATVGSAFYRD LSPGLGAWEAE KADGNVIRVG CAVCGEFFKA
 401 QVCEQLARKI EWLPSAQAL GIRNHYRHEP EHGSDRWANA LGSRRFSRNA
 451 CVVVSCTGAV TVDALDGDH YLGDTIMEGF HLMKESLAVR TANLNHAGK
 501 RYFPPTTTCN AVASGMMDAV CGSMMMHGR LKERTGAGKE VDVITGGGA
 551 AKVAELPPA FLAENTVRVA DNLVIYGLLN MIAAEGRYE HI*

Figure 9 shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF61-1. Further computer analysis of this amino acid sequence gave the following results:

25 Homology with the baf protein of *B. pertussis* (accession number U12020).

ORF61 and baf protein show 33% aa identity in 166aa overlap:

orf61 23 LLLDGGNSRLKWAUVE-NGTFATVGSAPYR----DLSPGLGAWEAEKADGNVIRVGCACVG 77
 +L+ GNSRLK W + + A AP DL LG A R +G V G
 baf 3 ILIDSGNSRLKVGWDFPDAPQAAREFAFVAFDNLDLALGRWLATLPRRQALGVNVAG 62
 orf61 78 EFKAQVQEQAR---KIEWLPSSAQAGIRNHYRHEPEHGSDRW---FNLGSRFRSRN 131
 + + L I WL + A G+RN YR+P+ G+DRW L +
 baf 63 LARGEIAATLRAGGCDIRWLRAQLAMGLRNGYRNPDLQADRWACMVGLARQPSVHP 122
 orf61 132 ACVVVSCTGAVTVDALTDGHHYLGXGTMIPGFHLMKESLAVRANTL 177
 +V S GTA T+D + D + G G I+PG +M+ +LA TA+L
 baf 123 PLLVASFGTATTDLTIGTGNVFPF-GLILPGFAMMRGALAYGTAHL 167

Homology with a predicted ORF from *N. meningitidis* (strain A)

40 ORF61 shows 97.4% identity over a 189aa overlap with an ORF (ORF61a) from strain A of *N. meningitidis*:

orf61.pep 10 20 30
 EISLRSDXRPVSVXKRRDSERFLLLDGGNS
 45 orf61a TVFEGTVKGVGDGGLHLETAEGKQTVVSSEI SLRSDDRPVSVFKRRDSERFLLLDGGNS
 290 300 310 320 330 340
 orf61.pep 40 50 60 70 80 90
 RLKWAUVVNGTFATVGSAPYRDLSPLGAWEAEKADGNVIRVGCACVGGEFFKAQVQEQAR
 50 orf61a RLKWAUVVNGTFATVGSAPYRDLSPLGAWEAEKADGNVIRVGCACVGGEFFKAQVQEQAR
 350 360 370 380 390 400
 orf61.pep 100 110 120 130 140 150
 KIEWLPSSAQAGIRNHYRHEPEHGSDRWFNALGSRFRSRNACVVVSCTGAVTVDALTD
 55 orf61a KIEWLPSSAQALGIRNHYRHEPEHGSDRWFNALGSRFRSRNACVVVSCTGAVTVDALTD
 410 420 430 440 450 460

```

      160      170      180      189
orf61.pep  GHYLGXGTIMPGFHLMKESLAVRTANLNHRHAGKRYFPFT
5 orf61a    ||||| ||||| ||||| ||||| ||||| ||||| |||||
      470      480      490      500      510      520
      GHYLG-GTIMPGFHLMKESLAVRTANLNHRHAGKRYFPFTTTGNVAVSGMMDAVCGSVMM
orf61a      HGRLKEKKTGAGKPVVDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIHGLNLLIAAEGG
      530      540      550      560      570      580

```

10 The complete length ORF61a nucleotide sequence <SEQ ID 235> is:

```

1 ATGACGGTTT TGAAGCCTTC GCACCTGGCG GTGTTGGCG AGCTTGCCGA
51 CGGTTTGGCG CAACACGTCT CGCAACTGGC GCGTATGGCG GATATGAAGC
101 CGCAGCAGCT CAACGGTTTT TGGCAGCAGA TGCCGGCGCA CATACGGCGG
151 CTGTTGCGCC AACACGACGG CTATTGGCGG CTGGTGCGCC CATTGGCGGT
15 201 TTTTCGATGCC GAAGGTTTGC GCGAGCTGGG GGAAGGCTGG GGTTTTCAGA
251 CGGCATTGAA GCAACGAGTGC GCGTCCAGCA ACGACGAGAT ACTGGAATTG
301 GCGCGGATTG GCGCGGACAA GCGCGACAAA ACCATTCTGT TGACCCACTT
351 GCAAAAGTAAG GGCAGSGGCG GGCAGSGGCG GAAGTGTGCG CACCGTTTGG
401 GCGAGTGTCT GATGTTCAST TTGTCGTCGG GTGTTGCGCA GTGGCGTGCC GCGCGCCTAT
20 451 GATTGCGGTT CGTGTGTCGCC TTGTCGTCGG GTGTTGCGCA GTGGCGTGCC GCGCGCCTAT
501 CTCGCGTTTG GTTTTGAAAA CGCAAACTCA GTGGCCAAAC GATTTGGTGC
551 TCGGACGCGA CAAATTTGGG GGCATTCTGA TTGAAACGGT CAGGACGGCGG
601 GGCAAAACGG TTGCGTGGT GCGTATCGGC ATCAATTTTC TGCTGCCCAA
25 651 GGAAGTGGAA AACGCGCTT CGTGCAATC GCTGTTTCAG ACGGCATCGC
701 GCGCGGGAAG TGCCGATGCC GCCGTGTTGC TGGAACCGCT GTTGGCGGAA
751 CTTGATGCGG GTTGTGTTGA ATATGCGCGG GACGGATTG CGCCTTTTGT
801 GCGCGAATAT CAGGCTGCCA ACCGCGACCA CGGCAAGCGG GTATTGCTGT
851 TGCGCGACGG GCAAAACGCTG TTCGAAGGCA CGGTAAAGG CGTGGACGGA
30 901 CAAGCGCTTC TGCACTTGGG AACGCGCAGG GGCACACAGA CGCTGCTCAG
951 CCGCGAATAT AGCCTGCGGT CGAAGCAGAC GCGCGTTTCC GTGCCGAAGC
1001 GCGCGGATTC GGAACGTTTT CTGCTGTTGG ACGCGGCGAA CAGCCGCTTC
1051 AAGTGGCGGT GTGTGGAAAA CGGACGCTTC GCAACGCTCG GACGCGCGC
1101 GTACCGCGAT TTGTCGCTCT TGGCGCGGAA GTGCGCGGAA AAGTGGATG
35 1201 CAAGTCCAGG AACAGCTCGC CGCAAAATC GAGTGGCTGC CGTCTTCGCG
1251 ACAGGCTTTG GGCATACGCA ACCACTACCG CCACCCCGAA GAACAGCGGT
1301 CCGACCGCTG GTTCAACGCC TTGGGACGCC GCGCGTTTCAG CCGCAACGCC
1351 TCGCTGCTCG TCAGTTGCGG CACGGCGGTA ACGGTTGACG CGCTCACCGA
40 1401 TGACGGACAT TATCTCGGGG GAACCATCAT GCCCGGTTTC CACCTGATGA
1451 AAGAATCGCT CGCCGTCGGA ACCGCCAACC TCAACCGGCA CGCCGTAAG
1501 CGTTATCTCT TCCGACACAC AACGGGCAAT GCGCGGTTTC CACCTGATGA
1551 GGATGCGGTT TCGCGCTCGG TTATGATGAT GCACGGGCGT TTGAAGAGAA
1601 AAACCGGGCG GGCAGACCTG GTCGATGTCA TCATTACCGG CGGCGCGCGG
45 1651 GCAAAAGTTG CCGAAGCCCT GCGCGCTGCA TTTTGGGCG AAAATACGCT
1701 GCGGTTGGCG GACAACCTCG TCATTACGCG GCTGCTGAAC CTGATTGCGG
1751 CCGAAGGCGG GGAATCGGAA CATACTTAA

```

This encodes a protein having amino acid sequence <SEQ ID 236>:

```

1 MTVLKPSHWR VLAEADLGLP QHVSOLARMA DMKPPQLNGF WQMPAHIRG
50 51 LLRQHDGYWR LVRLAVFDA EGLRELGRS GFQALKHEC ASSNDELLEL
101 ARIAPDKAKH TICVTHLQSK GRGRQGRKWS HRLGECLMFS FGWVDRBPY
151 ELGSLSPVAA VACRRALSR LKLTQIKWPN DUVVGRDKLG GILLETVRTG
201 GKTVAVVVIG INFVLPEVE NAASVQSLFQ TASRRGNADA AVLETLIAE
251 LDAVLLQYAR DGFAFPVAEY QANRNDHKA VLLLRDGETV FEFTVKGVGD
301 GQVHLHLETA EKQTVVSGEI SLRSDDRPVS VPKRRDSERF LLLDGNLSRL
55 351 KWAVWENGTF ATVGSAFYRD LSLPGAEWAE KVDGNVRIVG CAVCGEFKKA
401 QVQEQALARKI EWLPSAQAOL GIRNHYRHEP EHGSDRWENA LGSRRFSRNA
451 CVVVSCTGVA TVDALTDHGH YLGTTIMPFG HLMKESLAVR TANLNHRAGK
501 RYFPFTTTGN AVASGMDAV CGSVMMHGR LKERTGAGKP VDVIITGGGA
551 AKVAEALPPA FLAENTVRVA DNLVIHGLLN LIAEGGESE HT*

```

60 ORF61a and ORF61-1 show 98.5% identity in 591 aa overlap:

```

      10      20      30      40      50      60
orf61a.pep  MTVLKPSHWRVLAEADLGLPQHVSQQLARMADMKPPQLNGFWQMPAHIRGGLLRQHDGYWR
      ||||| ||||| ||||| ||||| ||||| |||||

```

	orf61-1	MTVLKLSHWRLAEADGLPQHSQQLARMADMKPQQQLNGFWQMPAHIRGLLRQHDGYWR	
		10 20 30 40 50 60	
5	orf61a.pep	LVRLAVFDAGLEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKHTICVTHLQSK	
	orf61-1	LVRLAVFDAGLEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKHTICVTHLQSK	
		70 80 90 100 110 120	
10	orf61a.pep	GRGRGGRKWSHRLGECIMFSFGWVFDPRQYELGSLSPVAACRRALSRLGLRGTQIKWPN	
	orf61-1	GRGRGGRKWSHRLGECIMFSFGWVFDPRQYELGSLSPVAACRRALSRLGLDVQIKWPN	
		130 140 150 160 170 180	
15	orf61a.pep	DLVVGDRKLGILLETVRTGGKTVAVVGIGINFLPKVEENAASVQSLQTASRRGNADA	
	orf61-1	DLVVGDRKLGILLETVRTGGKTVAVVGIGINFLPKVEENAASVQSLQTASRRGNADA	
		190 200 210 220 230 240	
20	orf61a.pep	AVLLETLLAELDAVLLQYARDGFAPFVAEYQAANRDHGAVALLLRDGETVFEGTVKGVDG	
	orf61-1	AVLLETLLAELDAVLLQYARDGFAPFVAEYQAANRDHGAVALLLRDGETVFEGTVKGVDG	
		250 260 270 280 290 300	
25	orf61a.pep	QGVHLHLETAEGKQTVVSGEISLSDRRPVSVPKRRDSEFLLDDGNSRLKWAUVENGTF	
	orf61-1	QGVHLHLETAEGKQTVVSGEISLSDRRPVSVPKRRDSEFLLDDGNSRLKWAUVENGTF	
		310 320 330 340 350 360	
30	orf61a.pep	ATVGSAPYRDLSPGAEWAEEKVDGNVRIYGCACVGEFKKAQVQELARKIEWLPSSAQAL	
	orf61-1	ATVGSAPYRDLSPGAEWAEEKVDGNVRIYGCACVGEFKKAQVQELARKIEWLPSSAQAL	
		370 380 390 400 410 420	
35	orf61a.pep	GIRNHYRHPPEEHGSDRWFNALGSRFRSRNACVVVSCGTAVTVDALTDGHHYLGGTIMPGF	
	orf61-1	GIRNHYRHPPEEHGSDRWFNALGSRFRSRNACVVVSCGTAVTVDALTDGHHYLGGTIMPGF	
		430 440 450 460 470 480	
40	orf61a.pep	HLMKESLAVRTANLNRHAGKRYFPFTTTGNVAVSGMMDAVCGSVMMHGRLEKKTGAGKP	
	orf61-1	HLMKESLAVRTANLNRHAGKRYFPFTTTGNVAVSGMMDAVCGSVMMHGRLEKKTGAGKP	
		490 500 510 520 530 540	
45	orf61a.pep	VDVITGGGAARVAEALPPAFLAENTVRVADNLVHGLNLIAAEGGESEHTX	
	orf61-1	VDVITGGGAARVAEALPPAFLAENTVRVADNLVYGLNLNMAAEGREYHIX	
		550 560 570 580 590	
50	orf61a.pep		
	orf61-1		
55	orf61a.pep		
	orf61-1		

Homology with a predicted ORF from *N.gonorrhoeae*

ORF61 shows 94.2% identity over a 189aa overlap with a predicted ORF (ORF61.ng) from *N.*

gonorrhoeae:

orf61.pep	EISLRSDXRPVSVVKRRDSEFLLDDGNS	30
orf61.ng	TVCETGVTKVGDGRGVLHLETAEGEQTVVSGEISLRPDNRVSVVPKRPDSEFLLLEGGNS	211

5	orf61.pep	RLKAWVENGTFATVGSAPYRDLSP LGAEWAEKADGNVRI VGCVCGEFKKAQVQEQLAR	90
	orf61.ng		271
10	orf61.pep	KIEWLPSSAQAKGIRNHYRHPEEHGSDRWFNALGSSRRFSRNACVVSCGTAVTVDALTTDD	150
	orf61.ng		331
15	orf61.pep	GHYLKGXTIMPGFHLMKESLAVRTANLN RHAGKRYFPPT	189
	orf61.ng		390

An ORF61ng nucleotide sequence <SEQ ID 237> was predicted to encode a protein having amino acid sequence <SEQ ID 238>:

15	1	MFSFGWAFDR	PQYELGSLSP	VAALACRRAL	GCLGLETQIK	WFNDLVVGRD
	51	KLGGILLETV	RAGGKTAVVV	GIGINFVLPK	EVENAASVQS	LFQTASRRNG
20	101	ADAVALLETL	LAELGAVLEQ	YAEGEFAPFL	NEYETLRNRD	GKAVLLLRDG
	151	ETVCEGTVKG	VDGRGVHLLE	TAEGETQVVS	GEISLRPNDR	SVSPVKRPDS
25	201	ERFLLLEGGN	SRLKAWVVEN	GTATVGSAP	YRDLSPFGAS	WAEKADGNVR
	251	IVGCAVCGES	KKAQVKEQLA	RKIEWLPSSA	QALGIPRNHYR	HPEEHGSDRW
30	301	FNALGSSRRFS	RNACVVVSCG	TAVTVDALTD	DGHYLGGTIM	PGFHLMKESL
	351	AVRIANLNRP	AGKRYFPPTT	TGNVASGMM	DAVCSIMMM	HGRLEKKNKA
35	401	GKPYDVITTG	GGAAKVAEAL	FEAFLAENTV	RVANLNVHNG	LNLILAAEAG
	451	ESEHR*				

Further analysis revealed the complete gonococcal DNA sequence <SEQ ID 239> to be:

25	1	ATGACGGT	TGAAGCCTTC	GCAITGGCGG	GTGTTGGCGG	AGCTTGCCGA
	51	CGGTTTCCCG	CAACACGAT	CGCAATTGGC	CGGTGAGCGG	GACATGAACG
30	101	CGCAGCAGCT	CAACGGTTTT	TGGCAGCAGA	TGCOCGCCGA	TATACCGCGG
	151	CTGTTGCCGC	AACACGACGG	CTATTGGCGC	CTGGTGCGCC	CCTTGGCGGT
35	201	TTTCGATGCC	GAAGGTTTGC	GCGATCTGGG	GGAAGGTCG	GTTTTCAGA
	251	CGGCATTGAA	GCACGATGTC	GCGTCCAGCA	ACGACGAGAT	ACTGGAATTG
40	301	GCGCGGATTG	GCCCGGACAA	GCGCGACAAA	ACCATATGCG	TGACCCCACT
	351	GCAAGTAAAG	GGCAGGGGGC	GGCAGGGGGC	GAACTGGTCG	CACCGTTTGG
45	401	CGGAGTGCCCT	GATGTTCACT	TTCGGCTGGG	GTTTTCAGCC	GCGCAGATAT
	451	GAGTTGGGTT	CGCTGTCCGC	TGTTGCGGCA	CTTGGCTGCC	GCGCGCGTTT
50	501	GCGGTGTTTG	GTTTGGGAAA	GCAAAATCAA	TGGGCCAAAC	GATTTGGTCG
	551	TGCGACGCGA	CAAACTTGGC	GCAATTTCTG	TTGAAACAGT	CAGCGCGGCG
55	601	GGTAAACCGG	TTGCCGTGGT	CGGTATCGGC	ATCAATTTCTG	TGCTTCCCAA
	651	GGAAAGTGGAA	AACGCGCGTT	CGGTGCAGTC	GCTGTTTCAG	ACGGCATCGC
60	701	GGCGGGGCAA	TGCCGATGCC	GCCGTATTGC	TGGAACAATT	GCTTGCGGAA
	751	TTGGGCGCGG	TGTTGGAACA	ATATTGCGGA	GAAAGGTTTC	CGCCATTTTT
65	801	AAATGAGTAT	GAAACGGCCA	ACCGCGACCA	CGGCAAGCGC	GTATTGCTGT
	851	TGCGCGACGG	CGAAACCGTG	TGCGAAGCGA	CGGTTAAAGG	CGTGAACGGA
70	901	CGAGGCGTTC	TGCACTTGGG	AACGGCGAag	ggcgaACAGa	cggtcgtagc
	951	cggcgaaATC	AGcctGcGgc	cgacacaacg	GTCGgttttc	gtgccgaagc
75	1001	ggcgcggtATC	GgaacgtTTT	tTGctgtgtg	aaggcgggaa	cagccgGCTC
	1051	AACTGGCGCT	GggtggAaaa	cggcacgctt	gcacacgctg	gcagcgcgCc
80	1101	gAACCGGAT	TTTTCGCGCT	TGGCGCGGA	GTGGCGGAA	AAGGCGGATG
	1151	GAAATGCTCG	CATCTCGCGT	TGCGCGCGT	CGGCGGATC	CAAAAGGCGA
85	1201	CAGTGAAGG	AACAGCTCGC	CCGAAATAATC	CGGCGGCGC	CGCTTCGCGC
	1251	ACAGGCTTTG	GGCATACGCA	ACCACTACCG	CCACCCCGAA	GAACACGGTT
90	1301	CGGACCGTTG	GTTCAACGCC	TTGGGCAGCC	CGCGCTTCAG	CGCAACGCCC
	1351	TGCGTCTCGT	TCAGTTGCGG	CACGGCGGTA	ACGGTTGACG	CGCTCAACGA
95	1401	TGACGGACAT	TATCTCGCGG	GAACCATCAT	GCCCGGCTTC	CACCTGATGA
	1451	AAGAATCGCT	CGCGTTCGCA	ACCGCCAACT	TCACCGCGCC	CGCGCGCAAA
100	1501	CGTTACCCCT	TCCCGACCA	AACGGGCAAC	GCGCTCGCAA	CGGCGCATAT
	1551	GGACGCGGTT	TGCGCTCGA	TAATGATGAT	GCACGGCGCT	TTGAAGAA
105	1601	AAAACGGCGC	GGGCAAGCCT	GTCGATGTCA	TCATTACCGC	CGGCGGCGCG
	1651	GCGAAAGTGC	CGGAAGCCCT	GCGCGCTGCA	TTTTTGGCGG	AAAATACCGT
110	1701	GCGCGTGGCG	GACACCTCTG	TCATCCACGG	GCTGCTGAAC	CTGATTGCGC
	1751	CCGAAGCGCG	GGAATCGGAA	CACGCTTA		

This corresponds to the amino acid sequence <SEQ ID 240; ORF61ng-1>:

1 MTVLKPSHWR VLAEALADGLP QHVSQALAREA DMKFPQNLNGF WQQMPAHIRG

5 51 LLRQHDGYWR LVRPLAVFDA EGLRDLGERS GFQTALKHEC ASSNDEILEL
 101 ARIAPDKAHK TICVTHLQSK GRGRQGRKWS HRLGECMF5 FGWAFDRPOY
 151 ELGSLSPVAA LACRRALGCL GLETKIKWPN DLVVGRDKLG GILLETVRAG
 201 GKTVAVVIGG INFVLPEKEE NAASVQSLFQ TASRRGNADA AVLLETLLAE
 251 LGAVLEQYAE EGFAFPLNEY ETANRDHGKA VLLLRDGETV CEGTVKGVDG
 301 RGVHLLETAE GEQTVSSEI SLRDPNRSVS VKRPDSERF LLLCGNSRL
 351 KQAWVENGTF ATVGSAFYRD LSPGAEWAE KADGNVIRG CAVCGSKKA
 401 QVKEQLARKI EWLPSAQAAL GIRNHYRHPE EHGSDRWFA LGSRRSPRNA
 451 CVVVSCTAV TVDALTDGHH YLGGTIMPGF HLMKESLAVR TANINRPPGK
 10 501 RYFPPTTGN AVASGMDAV CGSIMMMGR LKEKNAGKP VDVITGGGA
 551 AKVAEALPPA FLAENTVRVA DNLVIHGLLN LIAEGGESE HA*

ORF61ng-1 and ORF61-1 show 93.9% identity in 591 aa overlap:

15 orf61ng-1.pep MTLVKPSHWRVLAELADGLPQHVSQALAREADMKPQQLNGFWQOMPAHIRGLLRQHDGYWR 60
 orf61-1 MTLVKLSHWRVLAELADGLPQHVSQALARMADMKPQQLNGFWQOMPAHIRGLLRQHDGYWR 60
 20 orf61ng-1.pep LVRPLAVFDAEGLRDLGERSGFQTALKHECASSNDEILELARIAPDKAHKTCVTHLQSK 120
 orf61-1 LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTCVTHLQSK 120
 25 orf61ng-1.pep GRGRQGRKWSHRLGECMF5FGWAFDRPQYELGSLSPVAAACRRALGCLGLETKIKWPN 180
 orf61-1 GRGRQGRKWSHRLGECMF5FGWVDRPQYELGSLSPVAAVACRRALGCLGLDVQIKWPN 180
 30 orf61ng-1.pep DLVVGRDKLGGILLETVRAGGKTVAVVIGGINFVLPEKEVNAASVQSLFQ TASRRGNADA 240
 orf61-1 DLVVGRDKLGGILLETVRTGKTVAVVIGGINFVLPEKEVNAASVQSLFQ TASRRGNADA 240
 35 orf61ng-1.pep AVLLETLLAE LGAVLEQYAE EGFAFPLNEY ETANRDHGKAVLLLRDGETV CEGTVKGVDG 300
 orf61-1 AVLLETLLVELDAVLQYARDGFAPFVAEYQANRDHGKAVLLLRDGETV FEGTVKGVDG 300
 40 orf61ng-1.pep RGVHLLETAE GEQTVSSEI SLRDPNRSVS VKRPDSERF LLLCGNSRLKQAWVENGTF 360
 orf61-1 QGVHLLETAE GKTVSSEI SLASDDRPVS VKRPDSERF LLLCGNSRLKQAWVENGTF 360
 45 orf61ng-1.pep ATVGSAFYRDL SPLGAEWAE KADGNVIRG CAVCGESKKAQVKEQLARKIEWLPSAQAAL 420
 orf61-1 ATVGSAFYRDL SPLGAEWAE KADGNVIRG CAVCGE FKKAQVKEQLARKIEWLPSAQAAL 420
 50 orf61ng-1.pep GIRNHYRHP EHGSDRWFNALGSRFRNACVVVSCGTAVTVDALTDGHHYLGTTIMPGF 480
 orf61-1 GIRNHYRHP EHGSDRWFNALGSRFRNACVVVSCGTAVTVDALTDGHHYLGTTIMPGF 480
 55 orf61ng-1.pep HLMKESLAVRTANINRPPAGKRYFPPTTGN AVASGMDAV CGSIMMMHRLKEKNAGKP 540
 orf61-1 HLMKESLAVRTANINRHAGKRYFPPTTGN AVASGMDAV CGSVMMHRLKEKTGAGKP 540
 60 orf61ng-1.pep VDVITGGGA AKVAEALPPA FLAENTVRVADNLVIHGLNLIAAEGGESEHAX 593
 orf61-1 VDVITGGGA AKVAEALPPA FLAENTVRVADNLVIHGLNLIAAEGGESEHAX 593

Based on this analysis, including the homology with the baf protein of *B. pertussis* and the presence of a putative prokaryotic membrane lipoprotein lipid attachment site, it is predicted that these proteins from *N. meningitidis* and *N. gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 29

The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 241>:

1 ATGTTTACC AAATCCTTGC CTTGATTATC TGGAGCAGCT CGTTTATTGC
 51 CGCCAAATAT GTCTATGGCG GCATCGATCC CGCATTTGATG GTCCGCGTGC
 101 GCCTGCTAAT TGGCGCGCTG CTTGCACTGC CCGCTGCGCG CCGTCACTGC
 5 GGCAGAGATTG CGCGTGAGGA ATGGAAGCCG TTGCTGATTG TTGCTGCTGT
 201 CAACATATGTG CTGACCTGTC TGCTTCAGTT TGTCTGGGTTG AAATACACTT
 251 CGCGCCGCGG CGCATCGGTC ATTGTCGGAC TCGAGCGGCT CGTCACTGTTG
 301 TTTGTCGGAC ACTTTTCTTT CAACGACAAA GCGCGTGCCT ACCACTGGAT
 351 ATGCGCGCGG CGCGCATTTG CCGGTGTCGC GTCTGATGT GCGGCGGCTG
 401 CGGAAGAGGG CGCGGAAGTC GGCTGGTTTC GTGCGCTGCT GGTGTTGTGT
 10 GCGGCGCGCG GCTTTTGTGC CGCTATGGGT CCGACGCAAA GCTGATTGTC
 451 ACGCATCGGC GCACCGGCAT TCACATCTGT TTCATTGCC GCGCATCGT
 551 TGATGTGCTT GCGGTTTTCG CTTGCTTTGG CGCAAAGTTA TACCGTGAC
 601 TGGAGCGTCG GGATGTTATG TTGCTGCTG TATTGGGTT TGGGTCG...

This corresponds to the amino acid sequence <SEQ ID 242; ORF62>:

1 MFYQILALII WSSSFIAAKY VYGGIDPALM VGVRLIIAAL PALPACRRHV
 51 GKIPREEWKP LLIVSFVNYV LTLLQLQFVGL KYTSAASASV IVGLEPLIMV
 101 FVGHHFFNDK ARAYHWICGA AAFAGVALLM AGGAEEGGEV GWFGLLVLL
 151 AGAGFCAAMR PTQRLIARIG APAFTSVSIA AASLMCLPFS LALAQSIVTD
 201 WSVGMVLSLL YLGLGC...

20 Further work revealed the complete nucleotide sequence <SEQ ID 243>:

1 ATGTTTACC AAATCCTTGC CTTGATTATC TGGAGCAGCT CGTTTATTGC
 51 CGCCAAATAT GTCTATGGCG GCATCGATCC CGCATTTGATG GTCCGCGTGC
 101 GCCTGCTAAT TGGCGCGCTG CTTGCACTGC CCGCTGCGCG CCGTCACTGC
 25 GGCAGAGATTG CGCGTGAGGA ATGGAAGCCG TTGCTGATTG TTGCTGCTGT
 201 CAACATATGTG CTGACCTGTC TGCTTCAGTT TGTCTGGGTTG AAATACACTT
 251 CGCGCCGCGG CGCATCGGTC ATTGTCGGAC TCGAGCGGCT CGTCACTGTTG
 301 TTTGTCGGAC ACTTTTCTTT CAACGACAAA GCGCGTGCCT ACCACTGGAT
 351 ATGCGCGCGG CGCGCATTTG CCGGTGTCGC GTCTGATGT GCGGCGGCTG
 401 CGGAAGAGGG CGCGGAAGTC GGCTGGTTTC GTGCGCTGCT GGTGTTGTGT
 30 GCGGCGCGCG GCTTTTGTGC CGCTATGGGT CCGACGCAAA GCTGATTGTC
 501 ACGCATCGGC GCACCGGCAT TCACATCTGT TTCATTGCC GCGCATCGT
 551 TGATGTGCTT GCGGTTTTCG CTTGCTTTGG CGCAAAGTTA TACCGTGAC
 601 TGGAGCGTCG GATGCTGATG GTTCTGCTG TATTGGGTT TGGGTCGCG
 651 CTGCTACCGC TATTGCTGCT GGAACAAAGG GATGACCGCT GTTCTGSCA
 35 701 ATGTTTCCGG ACTGTTGATT TCGCTCGAAC CCGTCTGCG GTGCTGCTG
 751 GCGGTTTGA TTTTGGCGCA ACACCTGTGC CCGCTGTCG CTTTGGGCT
 801 GTTGTGCTC ATGCCCGCCA CTTTGGTTCG CGGCGCGCTG TCGCATCAA
 851 AATAA

This corresponds to the amino acid sequence <SEQ ID 244; ORF62-1>:

1 MFYQILALII WSSSFIAAKY VYGGIDPALM VGVRLIIAAL PALPACRRHV
 51 GKIPREEWKP LLIVSFVNYV LTLLQLQFVGL KYTSAASASV IVGLEPLIMV
 101 FVGHHFFNDK ARAYHWICGA AAFAGVALLM AGGAEEGGEV GWFGLLVLL
 151 AGAGFCAAMR PTQRLIARIG APAFTSVSIA AASLMCLPFS LALAQSIVTD
 201 WSVGMVLSLL YLGLGCWGYA YWLWNKQMSR VEVNVSGLLI SLEPVVGVLL
 45 AVLLIGLEHLS FVSALGVFVV IATLIVAGRL SHOK*

Computer analysis of this amino acid sequence gave the following results:

Homology with hypothetical transmembrane protein HI0976 of *H. influenzae* (accession number Q57147)

ORF62 and HI0976 show 50% aa identity in 114aa overlap:

Orf62 1 MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLIIAALPALPACRRHV 60
 M YQILALIIWSSS I K Y ++DP L+V VR R KI + K
 HI0976 1 MLYQILALIIWSSSLIVGKLTYSMDPVLVVQVRLIIAMIIVMPLFLRRWKKIDKPMRQ 60
 Orf62 61 LLIVSFVNYVLTLLQLQFVGLKYTSAASASVIVGLEPLIMV FVGHHFFNDKARAY 114
 L ++F NY LLQF+GLKYTSA+SA ++GLEPL+VVGHHFF K +
 HI0976 61 LNWLAFFNYTAVFLLQFVGLKYTSAASAVTMIGLEPLLVVFGHHFFKTKQNGF 114

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF62 shows 99.5% identity over a 216aa overlap with an ORF (ORF62a) from strain A of *N. meningitidis*:

5	orf62.pep	10 20 30 40 50 60	MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLLIAALPALPACRRHVKGKIPREEWKP
	orf62a	10 20 30 40 50 60	MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLLIAALPALPACRRHVKGKIPREEWKP
10	orf62.pep	70 80 90 100 110 120	LLIVSFNVYLTLLLQFVGLKYTSAASASVIVGLEPLLMVFVGHFFHNKARAYHWICGA
	orf62a	70 80 90 100 110 120	LLIVSFNVYLTLLLQFVGLKYTSAASASVIVGLEPLLMVFVGHFFHNKARAYHWICGA
15	orf62.pep	130 140 150 160 170 180	AAFAGVALLMAGGAEEGGEVGFGLLVLLAGAGFCAAMRPTQRLIARIGAPAFTSVSIA
	orf62a	130 140 150 160 170 180	AAFAGVALLMAGGAEEGGEVGFGLLVLLAGAGFCAAMRPTQRLIARIGAPAFTSVSIA
20	orf62.pep	190 200 210	AASLMCLPFSIALAQSYTVDWSVGMVLSLLYLGLGC
	orf62a	190 200 210	AASLMCLPFSIALAQSYTVDWSVGMVLSLLYLGVGCSWYAYLWNKMSRVPANVSGLLI
25	orf62a	220 230 240	SLEPVVGVLAVLILGEHLSFVSVLGVFVVIAATLVAGRLSHQKX
	orf62a	250 260 270 280	

The complete length ORF62a nucleotide sequence <SEQ ID 245> is:

1	ATGTTTACC	AAATCCTGC	CCTGATTATC	TGGAGCAGCT	CGTTTATTGC
51	CGCCAAATAT	GTCTATGGCG	GCATCGATCC	CGCATTGATG	GTCCGGCGTGC
101	GCCTGCTGAT	TGCTGCGCTG	CCTGCACTGC	CCGCTGCCG	CGCTCATGTG
151	GGCAAGATTC	CGCGTAGGGA	ATGGAAGCG	TTGCTGATTG	TGTCGTTTGT
201	CAACTATGTG	CTGACCTGTC	TACTTCAGTT	TGTCGGGTTG	AAATACACTT
251	CCGCCGCCAG	CGCATCGTGC	ATTGTCGGAG	TCGAGCCACT	GCTGATGGTG
301	TTTGTCCGAC	ACTTTTCTCT	CAACGACAAA	CGCGTGCCT	ACCACGGAT
351	ATCGCGCGCG	GCGGCATTTC	CCGCTGTCGC	GCTGCTGATG	GCGGCGGTTG
401	CGGAAGAGGG	CGGCGAAGTC	GGCTGGTTGC	GCTGCTGCTG	GGTGTGTTTG
451	CGCGGCGCGG	GCTTTTGTGC	CGCTATGGGT	CCGACGCAAA	GGCTGATTCG
501	ACGCATCGCG	GCACCGGCAT	TCACATCTGT	TTCCATTGTC	GGCCGATGCT
551	TGATGCGCTC	CTGCTGTTGC	CTGCTGTTGC	CGCAAGATTC	TACCGTTGAC
601	TGAGCGCTCG	GATGCTGATT	GTCGCTGCTG	TATTTGGGCG	TGGGGTGCAG
651	CTGGTACGCC	TATGGGCTGT	GGACACAGGG	GATGAGCGCT	GTTCCGCGCA
701	ACGTTTCGGG	ACTGTTGATT	TCGCTCGAAG	CCGCTGTCGG	CGTGTGCTGT
751	GCGGTTTGA	TTTGGGCGCA	ACACCTGTGC	CCGCTGTCGG	TCTTGGGCGT
801	GTTTGTGCTC	ATCGCGCGCA	CCTTGGTTGC	GCGCGGCTGT	TCGCATCAAA
851	AATAA				

This encodes a protein having amino acid sequence <SEQ ID 246>:

1	MFYQILALII	WSSSFIAAKY	VYGGIDPALM	VGVRLLIAAL	PALPACRRHV
51	GKIPREEWKP	LLIVSFNVY	LTLLQFVGL	KYTSASASV	IVGLEPLLMV
101	FVGHFFNDK	ARAYHWICGA	AAFAGVALLM	AGGAEEGGEV	GWFGCLLVLL
151	AGAGFCAAMR	PTQRLIARIG	APAFTSVSIA	AASLMCLPFS	LALAQSYTVD
201	WSVGMVLSLL	YLVGCGSWYA	YWLWNKMSR	VFANVSGLLI	SLEPVVGVL
251	AVLILGEHLS	FVSVLGVFVV	IATLVAGRL	SHQK*	

ORF62a and ORF62-1 show 98.9% identity in 284 aa overlap:

orf62a.pep	MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLLIAALPALPACRRHVKGKIPREEWKP	60
orf62-1	MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLLIAALPALPACRRHVKGKIPREEWKP	60

	orf62a.pep	LLIVSFVNYVLTLLQLFVGLKYSASAASVIVGLEPLLMVFVGHFFFNDRKARAYHWICGA	120
	orf62-1	LLIVSFVNYVLTLLQLFVGLKYSASAASVIVGLEPLLMVFVGHFFFNDRKARAYHWICGA	120
5	orf62a.pep	AAFAGVALLMAGGAEEGGVGFGLLVLLLAGAGFCAAMRPTQRLIARIGAPAFTSVSIA	180
	orf62-1	AAFAGVALLMAGGAEEGGVGFGLLVLLLAGAGFCAAMRPTQRLIARIGAPAFTSVSIA	180
10	orf62a.pep	AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGLGCGWYAYWLWKNKMSRVFANVSGLLI	240
	orf62-1	AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGLGCGWYAYWLWKNKMSRVFANVSGLLI	240
	orf62a.pep	SLEPVVGVLAVLILGEHLSFVSIVLGFVVVIAATLVAGRLSHQKX	285
15	orf62-1	SLEPVVGVLAVLILGEHLSFVSALGVFVVIAATLVAGRLSHQKX	285

Homology with a predicted ORF from *N.gonorrhoeae*

ORF62 shows 99.5% identity over a 216aa overlap with a predicted ORF (ORF62.ng) from *N.gonorrhoeae*:

20	orf62.pep	MFYQILALI IWSSSFIAAKYVYGGIDPALMVGVRLLIALPALPACRRHVVKIPREEWKPF	60
	orf62ng	MFYQILALI IWGSSSFIAAKYVYGGIDPALMVGVRLLIALPALPACRRHVVKIPREEWKPF	60
25	orf62.pep	LLIVSFVNYVLTLLQLFVGLKYSASAASVIVGLEPLLMVFVGHFFFNDRKARAYHWICGA	120
	orf62ng	LLIVSFVNYVLTLLQLFVGLKYSASAASVIVGLEPLLMVFVGHFFFNDRKARAYHWICGA	120
30	orf62.pep	AAFAGVALLMAGGAEEGGVGFGLLVLLLAGAGFCAAMRPTQRLIARIGAPAFTSVSIA	180
	orf62ng	AAFAGVALLMAGGAEEGGVGFGLLVLLLAGAGFCAAMRPTQRLIARIGAPAFTSVSIA	180
	orf62.pep	AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGLGCGWYAYWLWKNKMSRVFANVSGLLI	216
	orf62ng	AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGLGCGWYAYWLWKNKMSRVFANVSGLLI	240

The complete length ORF62ng nucleotide sequence <SEQ ID 247> is:

```

1  ATGTTTACC  AAATCCTTGC  CCTGATTATC  TGGGGCAGCT  CGTTTATTGC
51  CGCCAAATAT  GTCTATGGCG  GCATCGATCC  CGCATTGATG  GTCGGCGTGC
101  GCCTGCTGAT  TGCCGCGCTG  CCTGCATCTC  CGCGCTCCGC  CCGTCATCTG
151  GGCAAGATTG  CGCGTAGAGG  ATGGAAGCCG  TTGCTGATTG  TGTCGTTCTG
201  CAACATATGT  CTGACCCCTG  TGCTTCAGTT  TGTCGCGTTG  AAATACACTT
251  CCGCGCCGAG  CGCATCGCTG  ATTGTCGGAC  TCGAGCCGCT  GCTGATGGTG
301  TTTGTCGGAC  ACTTTTCTT  CAACAGCAAA  GCGCGTGCT  ACCACTGGAT
351  ATGCGGCGCG  GCGGCATTG  CCGGTGTGCG  GCTGCTGATG  CGGGCGCGTG
401  CGGAAGAGGG  CGGCGAAGTC  GGCTGGTTCG  GCTGCTGCT  GGTGTTGTTG
451  GCGGCGCGCG  GCTTTTGTGC  GCCTATGGCT  CCGACGCAAA  GGCATGATGC
501  CCGCATCGGC  GCACCGGCAT  TCACATCTGT  TTCCATTGCC  GCGCATGCTT
551  TGATGTGCGT  GCCTGTTGCG  CTTCCTTTG  CGCAAAATTA  TACCGTGGAC
601  TGAAGCGTCG  GCATGCTATT  GTCCGCTGTT  TATTGGGTT  TCGGGTCCGG
651  CTGGTACCGC  TATTGGCTGT  GGAACAAGGG  GATAGCCGCT  GTTCCTGCCA
701  ACGCGTCGGG  ACTGTTGATT  TCGCTCGAAC  CGTCTGCGG  GGTGCTGTTG
751  GCGGTTTGA  TTTTGGGCGA  ACATTTATCG  CCGGTGTCGG  CTTGGGCGCT
801  GTTTGTGCTG  ATCGCGCGCA  CTTTCCGCG  CGGCGGCTG  TCGCGCAGGG
851  ACGGCGCAAA  CGGCAATGCC  GTCTGA

```

This encodes a protein having amino acid sequence <SEQ ID 248>:

55	1	MEYQILALII	WGSSEFIAAKY	VYGGIDPALM	VGVRLIIAAL	PALPACRRHV
	51	GKIPREEWKP	LLIVSFVNYV	LTLQLFVGL	KYTSASAASV	IVGLEPLIMV
	101	FVGHFFFNDR	ARAYHWICGA	AAFAGVALLM	AGGAEEGGEV	GWFGCLLVLL
	151	AGAGFCAAMR	PTQRLIARIG	APAFTSVSIA	AASLMCLPFS	LALAQSYTVD
	201	WSVGMVLSLL	YLGCGWYAY	YWLWKNKMSR	VFANASGLLI	SLEPVVGVL
60	251	AVLILGEHLS	FVSALGVFVV	IATFAGRL	SRRDAQNGNA	V*

ORF62ng and ORF62-1 show 97.9% identity in 283 aa overlap:

		10	20	30	40	50	60
5	orf62ng.pep	MFYQILALIIWGSSFFIAAKYVYGGIDPALMVGVRLIIAALPALPACRRHVKGIPREEWKP					
	orf62-1	MFYQILALIIWSSSFFIAAKYVYGGIDPALMVGVRLIIAALPALPACRRHVKGIPREEWKP					
		10	20	30	40	50	60
10	orf62ng.pep	LLIVSFVNYVLTLLQFVGLKYTSAASASVIVGLEPLLMVFGHFFFNDKARAYHWICGA					
	orf62-1	LLIVSFVNYVLTLLQFVGLKYTSAASASVIVGLEPLLMVFGHFFFNDKARAYHWICGA					
		70	80	90	100	110	120
		70	80	90	100	110	120
15	orf62ng.pep	AAFAGVALLMAGGAEGGEVGFGLVLLAGAGFCAAMRPTQRLIARIGAPAFTSVSI					
	orf62-1	AAFAGVALLMAGGAEGGEVGFGLVLLAGAGFCAAMRPTQRLIARIGAPAFTSVSI					
		130	140	150	160	170	180
		130	140	150	160	170	180
20	orf62ng.pep	AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGLGCGWYAYLWNKMSRVANASGLLI					
	orf62-1	AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGLGCGWYAYLWNKMSRVANASGLLI					
		190	200	210	220	230	240
		190	200	210	220	230	240
25	orf62ng.pep	SLEPVGVLAVLILGEHLSPVSALGVFVVIATFAAGRLSRRDAQNGNAVX					
	orf62-1	SLEPVGVLAVLILGEHLSPVSALGVFVVIATFAAGRLSRRDAQNGNAVX					
		250	260	270	280	290	
		250	260	270	280		
30							

Furthermore, ORF62ng shows significant homology to a hypothetical *H. influenzae* protein:

```

sp|Q57147|Y976_HAEIN HYPOTHETICAL PROTEIN HI0976 >gi|1074589|pir||B64163
hypothetical protein HI0976 - Haemophilus influenzae (strain Rd KW20)
>gi|1574004 (U32778) hypothetical [Haemophilus influenzae] Length = 128
Score = 106 bits (262), Expect = 2e-22
Identities = 56/114 (49%), Positives = 68/114 (59%)

Query: 1 MFYQILALIIWGSSFFIAAKYVYGGIDPALMVGVRLIIAALPALPACRRHVKGIPREEWKP 60
      M YQILAL+IW SS I K Y +DP L+V VR R KI + K
Sbjct: 1 MLYQILALIIWSSSIVGKLTYSMMDFVLVVQVRLIIAMIVMPLFLRWKKIDKPMRKQ 60

Query: 61 LLIVSFVNYVLTLLQFVGLKYTSAASASVIVGLEPLLMVFGHFFFNDKARAY 114
      L ++F NY LLQF+GLKYTSA+SA ++GLEPLL+VFGHFF K +
Sbjct: 61 LWLAFENYTAFLQLQFVGLKYTSAASAVTMIGLEPLLNVFVGHFFFKQNGF 114

```

Based on this analysis, including the homology with the transmembrane protein of *H. influenzae* and the putative leader sequence and several transmembrane domains in the gonococcal protein, it is predicted that these proteins from *N. meningitidis* and *N. gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 30

The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 249>:

```

1 ATGCGCCGTT TTCTACCGAT CGCAGCCATA TGCAGCmGms TCCTGkkgTGA
51 sGGACTGACG GGGCAACCG GCAGCACCAG TCCTGCTGGCG GATTATTTCCT
101 GGTGAGATTGT TGCCTTCAGC GCAATGCTGC TGCTGCTGTT GTCCGCGCGTT
151 TTGGCACGTT ATGTCATATT GCTGTTGAAA GACAGCGCGC ACGCGCATTT
201 CGGTTCCGtA srtYGCCAAA gSGCCTGkks TGGG.TATGTT TACGCTGTTT

```

251 GCOGKACTGC COGGGCGTGT TCTGTTCCGC TTTCGCCAC AGTTCATCAA
 301 CGGACACGATT AATTCTGTGT TCGGCACGCA TACCACGAGC GCGCTTGAAC
 351 GCAGCCTCAA TTGTGACCAAG TCCGCATTGA ATTTGGCGCG AGACAAACGCC
 401 CTCGCGACAG CGGTCCCGGT GCAGATAGAC CTCATCGGCG CGGCTTCCCT
 451 GCGCCGGGAT ATGGCGAGGG TGCTGGAACA TTACGCCGGC AGCGTTTGTG
 501 CCCAGCTTGC CCTGTACAAY ksCGAAGCGC GCAAAATCGA AAAAAGCATC
 551 AACCOCGACA AGCTCGATCA GCGCTTTCCA GGTAAAGCGC GTTGGGAAAA
 601 AATCCaACGG GCGGGTTTCGG T'ACGGGATT' GGAAGCATA GCGCGGATAT
 651 TGTaCGCGCA GGGCTGGCTG TCGCGGGGTA CGCAwACGG CGCGATTATC
 701 GCCTTGTTTT TCCGTCAGCC GGTTCGCCAA GCGCTGGCAG AGGATGCGCT
 751 yTTAATCGAA AAGGCAAGGG CGAAATATGC TGAGTTGAGT TACAGCAAAA
 801 AAGGTTTGCA GACCTTTTTC CTGGCAACCC TGCTGATTGC CTCGCTGCTG
 851 TCGATTTTTC TTGCACTGGT CATGGCACTG TAITTCGCCC GCGCTTTCGT
 901 CGAACCCTGC CTATCGCTTG CCGAGSGGGC GAAGCGGGTG GCGCAAGGCG
 951 ATTTTCAGCCA GACGCGCCCC GTGTTGCGCA ACAGCAGATT CGGACCGTTG
 1001 ACCaKGTGTG TCAACCACAT GACCGAGCAG CTTTCCATCG CCAAGGATGC
 1051 AGACGAGCGC AACCGCGGCG GCGAGGAAGC CGCCAGGCAT TATCTTGAAT
 1101 GCGTGTGGA GGGGCTGACC ACGGCGGTGG TGGTGTGGA CGARCAAGGC
 1151 TGTCTGAAAA CTTCAACAA AGCGCGGGT ACC.

This corresponds to the amino acid sequence <SEQ ID 250; ORF64>:

1 MRRFLPIAAI CAXLXXGLT AATGSTSSLA DYFWIIVAFS AMLLLVLSAV
 51 LARYVILLIK DRRDGVFSX KAKXPXXMP TLVAXLPGVF LFPFGPAQFIN
 101 GTINSWFND THEALERSLN LSKSALINLAA DNALGNAPV QIDLIGASL
 151 PGDMGRVLEH YAGSGFAQLA LYNXASGKIE KSINPHKLDQ PFPKGARWEK
 201 IQRAGSVRDL ESIGGVLYAQ GWLSAGTHXG RDYALFERQP VFKGVAEDAV
 251 LIEKARAKYA ELSYSKKGLO TFFLATLLIA SLSIFLALV MALYFARRFV
 301 EPVLSLAEGA KAVAQGDFO TRPVLNDEF GRLXLFNHM TEQLSLAKDA
 351 DERNRRREEA ARHYLECVLE GLTPTGVVDF EQGCLKTENK AAGT..

Further work revealed the complete nucleotide sequence <SEQ ID 251>:

1 ATGGCGCGTT TTCTACCGAT CGCAGCCATA TGCGCCGTCG TCCTGTTGTA
 51 CGGACTGAGC GCGGCAACCG GCAGCACCAG TCGCTGGCGC GATTATTTCCT
 101 GGTGGATTGT TGCCTTCAGC GCAATGCTGC TGCTGGTGTG GTCCGCGGTT
 151 TTGGCAGCATT ATGTATATAT GCTGTTGAAA GACAGGCGCG ACAGCGTATT
 201 CGGTTGCGCAG AFTGCCAAAC GCTTTCTGG GATGTTTACG CTGTTTGCCG
 251 TACTGCCCGG CGTGTTCCTG TTCCGCGGTT TCCGACATTT CATCAACGCG
 301 ACGATTATTT CGTGTTCGCG CACGAGTACC CACAGCGCGC TTGAACGAG
 351 CCTCAATTGT AGCAAGCTCCG CATTGAATTT GCGCGCAGC AACGCCCTCG
 401 GCAACGCGGT CCCCCTGCAG ATAGACCTCA TCGCGCGCGC TTCCCTGCCC
 451 GGGGATATGG GCAGGGTGCT GGAACATTAC CGCGGACGCG GTTTTGCCCA
 501 GCTTGCCCTG TACAATGCGC CAAGCGGCAA AATCGCAAAA AGCATCAACC
 551 CGCAACAGCT CGATCAGCCG TTTCAGGTA AGCGCGGTG GGAaaaaATC
 601 CAACGGGCGG GTTCGCTCAG GGATTTGAAA AGCATAGGCG GCGTATTGTA
 651 CGCGCAGGGC TGGCTGTGCG CGGCTACGCA CAACGGGCGC GATTACGCTT
 701 TGTTTTTCCG T'ACGCGGTT' CCCAAGGCGC TGGCAGAGGA TGCCGTCTTA
 751 ATCGAAAAGG CAAGGCGGAA ATATGCTGAG TTGAGTTACA GCAaaaaaAGG
 801 TTTGAGACC TTTTCTCTGG CAACCTGCTG GATTGCTCG CTGCTGTGGA
 851 TTTTTCTTGC ACTGGTTCATG GCACTGTATT TCGCCGCGCG TTTGTCGAA
 901 CCGCTCCATT CGCTTGCGGA GGGGCGGAAG GCGGTGGCGC AAGCGGATTT
 951 CCGCAGACG CCCCCTGTGT TCGSCACAGA CAGGTTGCGA CCGTTGACCA
 1001 AGTTTATGCA CCACTGAAC CAGACGCTTT CCAATCGCAA AGAAGACGAA
 1051 GAGCGCAACC GCGCGCGCGA GGAAGCCGCC AGGCATTATC TTGAATGCGT
 1101 GTTGGAGGGG CTGACCACGG CGCTGCTGTT GTTGACGAA CAAGGCTGTC
 1151 TGAaaACCTT CAAACAAGCG GCGGAACAGA TTTTGGGATG GCGGCTTACC
 1201 CCGCTGTGGG CGAGCAGCCG GCACGGTTGG CACGGCGTTT CGCGCGACGA
 1251 GTCCCTGCTT GCCGAAGTGT TTGCGGCCAT CGGCGCGCGC GCGAGTACGG
 1301 ACAaaACGGT CCATGTGAAA TATGCGCGCG CGACGATGC CAAaaTCTGT
 1351 CTGGGCAAGG CAACCGTCTC GCCCGAAGAC AACGGCAAC GCGTGTGTAAT
 1401 GGTGATTGAC GACATCACCG TTTTGATACA CGCGCAaaaa GAAGCGCGGT
 1451 GGGGCGAAGT GCGGAAGCGG CTGGCACACG AATTCGCAA TCCGCTCAGC
 1501 CCAATCCAGC TTTCGCGCGA ACGGCTGGCG TGGAAATTGG GCGGGAAGCT
 1551 GGATGAGCAG GATGCGCGAAA TCCTGACGCG TTCGACCGAC ACCATCGTCA
 1601 AACAGGTGGC GGCATTGAAG GAAATGGTC GAAATTCGG CAATTATGCG
 1651 CGTTCCTCTT CGCTCAAAAT GGAaaATCAG GATTGGAAC CTTTATCGG
 1701 GAAATCGCTG GCAATCGCTC AAGCGGCTCG GTGCGCGGTT CCGGCGAGCG
 1751 TTGCGCGCGA ACGCTGACG GTGCGCGCGG ATACGACCG CATGCGGACG
 1801 GTGCTGCACA ATATTTTCAA AAATCGCGCC GAGCGCGCGG AAGAGCGCGA

1851 TGTGCCCGCAA GTCAGGGTAA AATCGGAAAC AGGGCAGGAC GGTCCGATTG
 1901 TCCTGACGGT TCCGACACAC GGCAGAGGGT TCGGCAGGGA AATGCTGCAC
 1951 AACGCTCTCG AGCCGTATGT AACGCACAAA CCGCGCGGGA CGGATTTGGG
 2001 TCTGCCTCTG GTGAAAAAAA TCATTGAAGA ACACGGCGGC CGCATCAGCC
 2051 TGAGCAATCA GGATCGGGT GCGCGTGTG TCAGATCAT CTTCGCCAAA
 2101 ACGGTAATTA CTTATCGGTA G

This corresponds to the amino acid sequence <SEQ ID 252; ORF64-1>:

1 MRRFLPIAAI CAVVLLYGLT AATGSTSSLA DYFWWIVAFS AMLLLVLSAV
 51 LARYVILLIK DRRDGVFGSQ TAKRLSGMFT LVAVLPGVFL PGVSAQFING
 101 TINSWFGNDT HEALERSLNL SKSALNLAAD NALGNVAVFO IDLIGAASLP
 151 GDMGRVLEHY AGSGFAQLAL YNAASGKIEK SINPHKLDQP FPGKARWEKI
 201 QRAGSVRDLE SIGGVLYAQG WLSAGTHNGR DYALFFRQPV PKGVAEDAVL
 251 IEKARAKYAE LSYSKKGLQT FFLATLLIAS LLSIFLALVM ALYFARRFVE
 301 PVLSLAEGAK AVAQGDFSQT RPVLNRNDEF RLTKLFNHEMT EQLSIAKEAD
 351 ERNRREEEAA RHYLECULEG LTTGVVVEDE QGCLRTFNKA AEQILGMLPT
 401 PLWGSSRHGW HGVSAAQSLI AEVFAAIGAA AGTDKPFVHVK YAAEDDAKIL
 451 LGKATVLPED NGNGVVMVID DITVLIHAQK EAAWGEVAKR LAHEIRNELT
 501 PIQLSAERLA WKLGKGLDEQ DAQILTRSTD TIVKQVAALK EMVAERFNYA
 551 RPSLKLKLENG DINALIGDVL ALYERQPCRF AELAGEEPLT VAADTTPMRQ
 601 VLENIFKNRA EAREADAVPE YVVKSETQD GRIVLTVCND KKGGRGRELH
 651 NAFEPYVTDK PAGTGLGLPV VKKIIIEHGG RISLSNQDAG GACVRIILPK
 701 TVRTYA*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF64 shows 92.6% identity over a 392aa overlap with an ORF (ORF64a) from strain A of *N. meningitidis*:

	10	20	30	40	50	60
orf64.pep	MRRFLPTAAICAXXIXXGLTAATGSTSSLA DYFWWIVAFSAML LVLVLSAVLARYVILLIK					
orf64a	MRRFLPTAAICAVVLLYGLTAAATGSTSSLA DYFWWIVAFSAML LVLVLSAVLARYVILLIK					
	10	20	30	40	50	60
orf64.pep	DRRDGVFGSXXAKXPPXXMFTLVAXLPGVFLFGFPAQFINGTINSWFGNDTHEALERSLN					
orf64a	DRRDGVFGSQIAKR-LSGMFTLVAVLPGVFLFGVSAQFINGTINSWFGNDTHEALERSLN					
	70	80	90	100	110	120
orf64.pep	DRRDGVFGSXXAKXPPXXMFTLVAXLPGVFLFGFPAQFINGTINSWFGNDTHEALERSLN					
orf64a	DRRDGVFGSQIAKR-LSGMFTLVAVLPGVFLFGVSAQFINGTINSWFGNDTHEALERSLN					
	130	140	150	160	170	180
orf64.pep	LSKSALNLAADNALGNVAVFOIDLIGAASLPGDMGRVLEHYAGSGFAQLALYNXASGKIE					
orf64a	LSKSALNLAADNALGNVAVFOIDLIGAASLPGDMGRVLEHYAGSGFAQLALYNXASGKIE					
	120	130	140	150	160	170
orf64.pep	KSINPHKLDQPPFGKARWEKIQRAGSVRDLESIGGVLYAQGWLXSGTHNGRXYALFFRQP					
orf64a	KSINPHKLDQPPFGKARWEKIQRAGSVRDLESIGGVLYAQGWLXSGTHNGRXYALFFRQP					
	180	190	200	210	220	230
orf64.pep	KSINPHKLDQPPFGKARWEKIQRAGSVRDLESIGGVLYAQGWLXSGTHNGRXYALFFRQP					
orf64a	KSINPHKLDQPPFGKARWEKIQRAGSVRDLESIGGVLYAQGWLXSGTHNGRXYALFFRQP					
	240	250	260	270	280	290
orf64.pep	VPGKVAEDAVLIEKARAKYAE LSYSKKGLQTFFLATLLIASLLSFLALVMALYFARRFV					
orf64a	VPGKVAEDAVLIEKARAKYAE LSYSKKGLQTFFLATLLIASLLSFLALVMALYFARRFV					
	300	310	320	330	340	350
orf64.pep	EPVLSLAEGAKAVAQGDFSQTRPVLNRNDEFGRILTKLFNHEMTQLSIAKEADERNRREEA					
orf64a	EPVLSLAEGAKAVAQGDFSQTRPVLNRNDEFGRILTKLFNHEMTQLSIAKEADERNRREEA					

-197-

```

      370      380      390
orf64. pep  ARHYLCVLEGLTTGVVVVFDEQCLKTFNKAAGT
             |||||
orf64a     ARHYLCVLEGLTTGVVVVFDEQCLKTFNKAAEQILGMPLTFLWGSRRHGHGVSAAQQSL
5          360      370      380      390      400      410
orf64a     LAEVFAAIGAAAGTDKPFVHVKYAAPDDAKILLGKATVLPEDNXNGVMVIDDITVLIHAQ
             420      430      440      450      460      470

```

The complete length ORF64a nucleotide sequence <SEQ ID 253> is:

```

10      1  ATGCGCCGTT  TTCTACCGAT  CGCAGCCATA  TGCGCCGTCG  TCCTGTTGTA
      51  CGGACTGACG  GCGGCAACCG  GCAGCACCAG  TTCGTGGCGG  GATTATTTC
     101  GGTGGATTGT  TCGCTTCAGC  GCAATGCTGC  TGCTGGTGTT  GTCGCGCGTT
     151  TTGGCAGGTT  ATGTCAATTT  CGTGTGAAA  GACAGCGCGG  ACGGCGTATT
     201  CGGTTCGCAG  ATTGCCAAAC  GCCTTTCGG  GATGTTTACG  CTGGTTCGGC
     251  TACTGCCCGG  CGTGTTCTCG  TTCCGCGTTT  CCGCACAGTT  TATCAACGGC
     301  ACGATTAAAT  CGTGTTCTCG  CAACGATACC  CACGAGCGCG  TTGAACCGAG
     351  CCTCAATTTG  AGCAAGTCCG  CATTGAATCT  GCGGCGACAG  AACGCCCTTC
     401  GCAAGCCGAT  CCCCCTGCAG  ATAGACNTCA  TCGGCGCGCG  TTCGCTGGCC
     451  NSGGTATATG  GCAGGTGCTG  GGAACATATC  GCGCGGACGG  GTTTGCCGCA
     501  CCTTGCCCTG  TACAATGCCA  CAAGCGGCAA  AATCGACAAA  AGCATCAACC
     551  CGCACAAGCT  CGATCAGCCG  TTTCCAGGTA  AGCGCGCTTG  GGAATAAATC
     601  CAACAGCGCG  GTTCGCTCAG  GGATNNGGAA  AGCATAGCGG  GCGTATTGTA
     651  CGCGCANGCG  TGGCTGTCGG  CAGNNACGCA  CAACGGGCGC  GATTAGCGCT
     701  TGTTTTTTCG  TCAGCCGCTT  CCAAAAGCG  TGGCAGAGGA  TGCCGCTCTTA
     751  ATCGAAAAGG  CAAGGGCGNA  ANANNNTNAG  TTGAGTTACA  GCAAAAAAGG
     801  TTTTGAGACC  TTTTTCCTNG  CAACCCCTGT  ATTGCTCTCN  CTGCTGTCGA
     851  TTTTCTCTGC  ACTGGTCATG  GCACCTGATT  TCCGCCGCGC  TTTGCTCGAA
     901  CCGCTCCTAT  CGCTTCCGGA  GGGGCGAAG  CGGTGGCGC  AAGCGGATTT
     951  CAGCCAGACG  CGCCCGCTGT  TCGCGCAACA  CGAGTTCCGA  CGCTTGACCA
    1001  AGTTGTTCAA  CCACATGACC  GAGCAGCTTT  CATCTGCCAA  AGAAGCAGAC
    1051  GAGCGCAACC  CGCGGCGCGA  GGAAGCGCGT  AGACATTATC  TCGAATTCGT
    1101  GTTGGAGGGG  CTGACCACGG  CGGTGGTGCT  GTTGAGTGAA  CAGGCTCTTC
    1151  TGAARAAGCT  CAGACACAGA  CGCGACACGA  TTTTGGGATG  GCGCCCTCAT
    1201  CCGCTCTCGG  CGACGAGCGG  GCACGCTTGG  CAOGGCGTTT  CGCGCGACGA
    1251  GTCCCTGCTT  CGCGAAGTGT  TTGCGCCAT  CGCGCGCGCG  CGAGTTCACG
    1301  ACAAAACGGT  CCATGTGAAA  TATGCCGCGC  CGGACGATGC  CAAAATCTCT
    1351  CTGGGCAAGG  CAACCGTCTT  GCCCGAAGAC  AACNGCAACG  GCGTGTAAT
    1401  GGTGATTGAC  GACATCACCG  TTTTGATACA  CGCGCAAAA  GAAGCGCGCT
    1451  GGGGCGAAGT  GGCARAAACG  CTGGCACACG  AAATCGCAAA  TCCGCTCACG
    1501  CCCATCCAGC  TTTCTGCGGA  ACGGCTGGCG  TGGAAATTGG  GCGGGAGCTG
    1551  GGACGAGCAN  GACGCGCAAA  TCCTGACACG  TTCGACCGAC  ACCATCATCA
    1601  AACCAAGTGC  GGCATTAAAA  GAAATGGTCG  AGGCATTCCG  CAATTACNCG
    1651  CGTTCGCCCT  GNCNCTCAAT  GAAAATCAG  GATTGAACG  CCTTAATCGG
    1701  CGATGTGTTG  GCATTGTACG  AAGCTGGTCC  GTGCCGGTTT  GCGGCGGAAC
    1751  TTGCGCGCGA  ACCGCTGATG  ATGGCGCGCG  ATACGACCGC  CATCGCGCAG
    1801  GTGCTGCACA  ATATTTTCAA  AAATGCGCCG  GAAGCGCGCG  AAGAAGCCGA
    1851  TGTGCCCCAA  GTCAGGCTAA  AATCGGAAGC  GGGCGAGGAC  GAGCGGATAT
    1901  TCTTCGACAG  TTGCAACAC  GCGAGGGGT  TCGCAGAGGA  AATGCTGCAC
    1951  AATCGCTTCG  AGCCGTATGT  AACGACAAA  CCGGCTGAAA  CGGATTTGNG
    2001  ACTGCCCTGT  GTGAAAAAAA  TCATTGAAGA  ACACGCGCGC  CNCATCAGCG
    2051  TGAGCAATCA  GATATCGGGC  GCGCCGTNTG  TCAGATCATC  CTTGCCAAAA
    2101  ACGGTAGAAA  CTTATCGCTA  G

```

This encodes a protein having amino acid sequence <SEQ ID 254>:

```

55      1  MRRFLPIAAI  CAVVLLYGLT  AATGSTSSLA  DYFWWIVAFS  AMLLLVLSAV
      51  LARYVILLKK  DRRDGVFSGQ  TAKRLSGMFT  LVAVLPGVFL  FGVSAQFING
     101  TINSWFGNDT  HEALERSLNL  KSKALNLAAD  NALGNAPVQ  IDXIGASLP
     151  XDMGRVLEHY  AGSGFQAQL  YNAASKIEK  SINPHKLDQP  FPGKARWEKI
     201  QQAGSVRDXE  SIGGVLYAXG  WLSAXTHNGR  DYALFFRPQV  PKGVAEDAVL
     251  IERAKARXXX  LSYSKKGLQT  FFLATLLIAS  LLSIFLALVM  ALYFARRFVE
     301  PVLSLAEGAK  AVAQGDFSTQ  RFLVRNDEFG  RLTKLFNHMT  EQLSLAKEAD
     351  ERNRKREAAA  RHLYLCVLEGL  TTVGVVVFDE  QCCLKTFNKA  ABQLLMPLTT
     401  FLWGSRRHGH  YAAFDKLLR  REVFALYAGA  LKATGCTGCT  YAAFDKLLR
     451  LGKATVLPED  NXNGVMVIDD  DITVLIHAQK  EAAWGEVAKE  LAHRINPLIT
     501  PIQLSARLER  WKLGKGLDEX  DAQILTRSTD  TTIKQVAALK  EMVEAFPNRY
     551  RSPSKQLENQ  DLNALIGDVL  ALYEAGPCRF  ABEALAGEFLM  MAADTTAMRK
     601  VLHNIFKNAA  EAAEADVPE  VRVKSEAGQD  GRIVLTVCDN  KGFGGREMLH

```

651 NAFEPYVTDK PAGTGLKLPV VKKIIIEHGG XISLSNQDAG GAXVRIILPK
701 TVETYA*

ORF64a and ORF64-1 show 96.6% identity in 706 aa overlap:

		10	20	30	40	50	60
5	orf64a.pep	MRRFLPIAAICAVVLLYGLTAATG	TSSSLADYFWWIVAFS	AMLLLVLSAVLARYVILLK			
	orf64-1	MRRFLPIAAICAVVLLYGLTAATG	TSSSLADYFWWIVAFS	AMLLLVLSAVLARYVILLK			
		10	20	30	40	50	60
10	orf64a.pep	DRRDGVFGSQIAKRLSGMFTLVAVLP	GVFLFGVSAQFINGTINSWFGND	THEALERSLNL			
	orf64-1	DRRDGVFGSQIAKRLSGMFTLVAVLP	GVFLFGVSAQFINGTINSWFGND	THEALERSLNL			
		70	80	90	100	110	120
15	orf64a.pep	SKSALNLAADNALGNAIPVQID	XIGAASLPKDMGRVLEHYAGSG	FAQLALYNAASGKLEK			
	orf64-1	SKSALNLAADNALGNAIPVQID	XIGAASLPKDMGRVLEHYAGSG	FAQLALYNAASGKLEK			
		130	140	150	160	170	180
20	orf64a.pep	SINPHKLDQPPFGKARWEKIQAGS	VRDLESIGGVLYAGWL	SAXTHNGRDYALFRQPV			
	orf64-1	SINPHKLDQPPFGKARWEKIQAGS	VRDLESIGGVLYAGWL	SAXTHNGRDYALFRQPV			
		190	200	210	220	230	240
25	orf64a.pep	PKGVAEDAVLIEKARAXXXLSYS	SKGQLQTFFLATLLIASLLSIF	LALVMALYFARRVE			
	orf64-1	PKGVAEDAVLIEKARAXXXLSYS	SKGQLQTFFLATLLIASLLSIF	LALVMALYFARRVE			
		250	260	270	280	290	300
30	orf64a.pep	PVLSLAEGAKAVAQGSQTRPVL	RNDEFGRLTKLFNHMTQ	QLSIAKEADERNRRREFAA			
	orf64-1	PVLSLAEGAKAVAQGSQTRPVL	RNDEFGRLTKLFNHMTQ	QLSIAKEADERNRRREFAA			
		310	320	330	340	350	360
35	orf64a.pep	RHYLECVLEGLTTGVVVFDEQ	GCCLKTFNKAAEQILGMPLT	PLWGSSRHGWGHVSAQQSL			
	orf64-1	RHYLECVLEGLTTGVVVFDEQ	GCCLKTFNKAAEQILGMPLT	PLWGSSRHGWGHVSAQQSL			
		370	380	390	400	410	420
40	orf64a.pep	AEVFAAIGAAAGTDKPVHV	KYAAPDDAKILLGKATV	LPEDNMGVVMVDDITVLI	HAQK		
	orf64-1	AEVFAAIGAAAGTDKPVHV	KYAAPDDAKILLGKATV	LPEDNMGVVMVDDITVLI	HAQK		
		430	440	450	460	470	480
45	orf64a.pep	EAAWGEVAKRLAHEIRNPL	PIQLSAERLAWKLGGL	DEXDAIILTRSTDTIIKQ	VAALK		
	orf64-1	EAAWGEVAKRLAHEIRNPL	PIQLSAERLAWKLGGL	DEXDAIILTRSTDTIIKQ	VAALK		
		490	500	510	520	530	540
50	orf64a.pep	EMVEAFRNYXRSFSXQ	LENQDINALIGDVLAL	YEAAGPCRF	FAELAGEPLMMAADTTAM	RQ	
	orf64-1	EMVEAFRNYXRSFSXQ	LENQDINALIGDVLAL	YEAAGPCRF	FAELAGEPLMMAADTTAM	RQ	
		550	560	570	580	590	600
55	orf64a.pep	VLHNIFKNAEAAEEADVP	EVVRKSEAGQDGRIVL	TVCDNGKGFGRMLHNA	FEPPYVTDK		
	orf64-1	VLHNIFKNAEAAEEADVP	EVVRKSEAGQDGRIVL	TVCDNGKGFGRMLHNA	FEPPYVTDK		
		610	620	630	640	650	660
60	orf64a.pep	VLHNIFKNAEAAEEADVP	EVVRKSEAGQDGRIVL	TVCDNGKGFGRMLHNA	FEPPYVTDK		
	orf64-1	VLHNIFKNAEAAEEADVP	EVVRKSEAGQDGRIVL	TVCDNGKGFGRMLHNA	FEPPYVTDK		
		610	620	630	640	650	660

		670	680	690	700
orf64a.pep	PAGTGLXLPVVKKIIIEHGGXISLSNQDAGGAXVRIILPKTVETYAK				
orf64-1	PAGTGLGLFPVVKKIIIEHGGGRISLSNQDAGGACVRIILPKTVKTYAK				
		670	680	690	700

Homology with a predicted ORF from *N.gonorrhoeae*

ORF64 shows 86.6% identity over a 387aa overlap with a predicted ORF (ORF64.ng) from *N.*

10 *gonorrhoeae*:

orf64.pep	MRRFLPIAICAXXKXGLTAATGSTSSSLADYFWWIVAFSAMLILLVLSAVLARYVILLKK	60
orf64.ng	MRRFLPIAICAVVLLYGLTAATGSTSSSLADYFWWIVSFSAMLILLVLSAVLARYVILLKK	60
orf64.pep	DRRNGVFGSXXAKXPXXMFTLVAXLPGVFLFGFPQAQINGTINSWFGNDTHEALERSLN	120
orf64.ng	DRRNGVFGSQIAKR-LSGMFTLVAVLPGFLFPGISAQINGTINSWFGNDTHEALERSLN	119
orf64.pep	LSKSALNLAADNALGNAPVQIDLIGAASLPGDMGRVLEHYAGSGFAQLALYNXASGKIE	180
orf64.ng	LSKSALDLAADVSNAPVQIDLTASLGGNMGSVLEHYAGSGFAQLALYNAASGKIE	179
orf64.pep	KSINPHKLDQFPFGKARWEKIQRAGSVRDLESIGGVLYAQGWSAGTHXGRDYALFFRQP	240
orf64.ng	KSINPHQFDQPLDKEHWEQIQQTGSVRSLESIGGVLYAQGWSAGTHNGRDYALFFRQP	239
orf64.pep	VPKGVAEDAVLIEKARAKYAEISYSKKGLOTFFLATLLIASLLSIFLALVMALYFARRFV	300
orf64.ng	IPENVAQDVLIEKARAKYAEISYSKKGLOTFFLVTLIASLLSIFLALVMALYFARRFV	299
orf64.pep	EPVLSLAEGAKAVAQGDFSQTRFVLNRNDEFGRLLTXLFNHMTQLSLIAKADERNRRREEA	360
orf64.ng	EPILSLAEGAKAVAQGDFSQTRFVLNRNDEFGRLLTXLFNHMTQLSLIAKADERNRRREEA	359
orf64.pep	ARHYLECVLEGLTTGVVVDFEQGLKTFNKAAGT	394
orf64.ng	ARHYLECVLDGLTTGVVVSYP LSCCRATVFTSCHSSPLSYF	400

An ORF64ng nucleotide sequence <SEQ ID 255> was predicted to encode a protein having amino acid sequence <SEQ ID 256>:

40	1	MRRFLPIAAT	CAVVLLYGLT	AATGSTSSLA	DYFWWIVSFS	AMILLVLSAV
	51	LARYVILLKK	DRRNGVFGSQ	IAKRSLSGMFT	LVAVLPGFLF	FGISAQFING
	101	TINSWFGNDT	HEALERSLNL	SKSALDLAAD	NAVSNAPVQ	IDLITGASLS
	151	GNMGSVLEHY	AGSGFAQLAL	YNAASGKIEK	SINPHQFDQ	LPDKEHWEQ
	201	QQTGSVRSLE	SIGGVLYAQG	WLSAGHNHR	DYALFFRQPT	PEWVAQDVL
45	251	IEKARAKYAE	LSYSKKGLO	TFLLIASLLS	IFLALVMALY	ALYFARRFV
	301	PILSLAEGAK	AVAQGDFSQT	RPVLRNDEF	GRLLTXLFNHMT	QLSLIAKEAD
	351	ERNRRREEA	RHYLECVLDG	LTTGVVVSYP	LSCCRATVFS	TCHSSPLSYF*

Further work revealed the complete gonococcal DNA sequence <SEQ ID 257>:

	1	ATGCGCGCCT	TCCTACCGAT	CGCAGCCATA	TGCGCGCTCG	TCCTGCTGTA
	51	CGGATTGACG	CGCGCGACCG	CGACGACCAG	TTCGCTGGCG	GATTATTCTT
	101	GGTGGATAGT	CTCGTTACAG	CGAATGCTGC	TGCTGGTGTT	GTCGCGCGTT
	151	TTGCGCAGTT	ATGTCATATT	GCTGTTGAAA	GACAGGCGCA	ACGCGCTGTT
	201	CGGTTCCGAG	ATTGCCAAAC	GCCTTTCGGG	GATGTTCCAG	CTGGTCGCGC
	251	TACTGCCCGG	CTTGTTCTCT	TTCGGCATTT	CCGGCGAGTT	TATCAACGCG
55	301	ACGATTAAAT	CGTGGTTCGG	CAACGACACC	CACGAAGCCC	TGAAACGAG
	351	CCTTAAATTT	AGCAAGTCCG	CACCTGGATT	GGCGGCAGAC	AATGCCGTCG
	401	GCAACGCGCT	TCCCGTACAG	ATAGACCTCA	TCGGCACCGC	CTCCGTGTCG
	451	GGCAATATGG	GCAGTGTGCT	GGACACTAC	GCCGCGACGG	GTTTTCGCCA
	501	GCTTGCCTGG	TACATGTCAG	CAATCGAARA	AGCATCATAC	AGCATCATAC
60	551	CGCACCATTT	CGACGACCGG	CTTCCGACGA	AAGAACATTG	GGAACAGATT

601 CAGCAGACCG GTTCGGTTCG GAGTTTGGAA AGCATAGCGG GCGTATTGTA
 651 CGCGCAGGGA TGGTTGTCGG CAGGTACGCA CAACGGGCGG GATTACGGCG
 701 TGTTCCTCCG CAGCCGATTT CCGCAAAATG TGCCACAGGA TCGCGTCTCG
 751 ATTGAAAAAG CGCGGGCGAA ATATGCCGAA TTGAGTTACA GCIAAAAAAGG
 801 TTTGCAGACC TTTTTCCTGG TAACCTTGCT GATTGCCTCG CTGCTGCGGA
 851 TTTTTCCTGC GCTGGTAATG GCATCTGATT TTGCCCCGCG TTTCTGTCGAA
 901 CCCATTCTGT CGCTTGCCGA GGGCGCAAGG GCGGTGCGCG AGGGTGATT
 951 CAGCCAGACG CGCCCCGTAT TGCGCCAACA CGAGTTCGGA CGTTTGACCA
 1001 AGCTGTTCAA CCATATGACC GAGCAGCTTT CCATGCCCAA AGAAGCAGAC
 1051 GAACGCAACC GCCGGCGCGA GGAAGCGCGC CGTCACTACC TCGAGTCGGT
 1101 GTTGGATGGG TTGACTACCG GTGTGGTGGT CATTGACGAA AAGGCGCGTT
 1151 TGAACACCTT CAACAGGCGG CGCGAACAGA TTTTGGGGAT GCCGCTCGCC
 1201 CCCCTGTGGG GCAGCAGCGG GCACGGTTTG CACGGCGTTT CGCGCGCAGA
 1251 GTCCCTGCTT GCCGAAGTGT TtgccgcCAT CGGTGCGCGG CGAGGTACGG
 1301 ACAACACCGT CAGGTTGGAA TATGCCGCGC CGGACGATGC CAAATCTCGT
 1351 CTGGGCAAGG CGACGGTATT GCCCGAAGAC AACCGCAAGC GCCTGGTGAT
 1401 GGTGATTGAC GACATCACCG TGCTGATACG CGCGCAAAAG GACCGCCGCT
 1451 GGGGTGAAGT GCGCAAGCGG CTGGCAGACG AATTCGCGAA TCGCTCACG
 1501 CCCATCCAGC TTTTCGCGGA ACCTGTGGCG TGGAAATGCG GCGGGAAGCT
 1551 GAGCACTCAG GACGCGCAAA TCCTGACGCG TtcgACGAC ACCATCACTA
 1601 AACAGTgtgc gGCGTTAAAA GAAATGGTGC AGGCATTCCG CAATTACCGG
 1651 CGCGCCCTCT CGCTCAAAT GGGAAATCAG GATTGAACGC CTTTAATCGG
 1701 CGATGTTTTC GCCCTGTACG AAGCCGCGCC GTGCGGGTTT GAGGCGGAAAC
 1751 TTGCGCGCGA ACCGCTGATG ATGGCGCGGG ATACGACCGC CATGCGCGAG
 1801 GTGCTGCACA ATATTTTCAA AAATCGCCCG GAAGCGCGGG AAGAAGCCGA
 1851 TATGCCCGAA GTCAGGGTAA AATCGGAAAC GGGCAGGACG GACGCGATTG
 1901 TCCTGACGGT TTGCGACAAC GGCAGGGAT TCGGCAAGGA AATGCTGCAC
 1951 AATGCTTTTC AGCCGTATGT GACGGATAAG CCGCGGGGAA CGGACTCTGG
 2001 TCTGCTGTGA GTGAAAAAAA TCATTGGAGA ACACGCGCGC CGCATCAGCC
 2051 TGAGCAATCA GGATGCGGGT GGGGCGTGTG TCAGAATCAT CTGCAAAAAA
 2101 ACGGTAGAAA CTTATGCGTA G

This corresponds to the amino acid sequence <SEQ ID 258; ORF64ng-1>:

1 MRRFLPTAAI CAVVLLYLGLT AATGTSLSLA DYFWWVSFSF AMLLLVLSAV
 51 LARYVILLLK DRNRGVFSGQ LARKLSGMFT LVAVLPGLFL FGIQAQFNG
 101 TINSWFGNDT HEALERSNL SKSALDLAAD NAVSNAPVQ TDLIGTASLS
 151 GNMGSVLEHY AGSGFAQLAL YNAASGKIEK SINPHFPQDP LDPKEHWQPI
 201 QQTGSVRSLE SIGGVLYAQG WLSAGTHNGR DYALFPRQPI PENVAQDVAL
 251 IEKARAKYAE LSYSKKQLQT FFLVTLIIAS LLSFLALVM ALYFARRFVE
 301 PILSLAEGAK AVAQGDFSQT RFLVRNDEFG RLTKLFNHMT EQLSIAKEAD
 351 ERNRRRREEA RHYLECVDLG LTTGVVVFDE KGRILKTFNKA AEQLLGMPLA
 401 PLWGSRRHGW HGVSAAQSL L AEFVFAIGA AGTDPKVQVE YAAPDDAKIL
 451 LGKATVLPED NGNGVVMVID DITVLIRAQK EAAWGEVAKR LAHEIRNPLT
 501 PIQLSAERLA WKLGKLDQDQ DAQILTRSTD TIIKQVAALK EMVEAFRNYA
 551 RAPSLKLENQ DLNALIGDVL ALYEAGPCRF EAEALAGEPLM MAADTTAMRQ
 601 VLHNIFKNAA EAAEADMPPE VRVKSETQGD GRIVLVCDN KGKFGKEMHL
 651 NAFEPYVTDK PAGTGLGLPV VKKIIEHGEG RISLSNQDAG GACVRIILPK
 701 TVETYA*

ORF64ng-1 and ORF64-1 show 93.8% identity in 706 aa overlap:

		10	20	30	40	50	60
50	orf64ng-1.pep	MRRFLPTAAICAVVLLYLGLTAA	TGTSLSLADYFWWVSFSAMLLLVLSAVLARYVILLIK				
	orf64-1	MRRFLPTAAICAVVLLYLGLTAA	TGTSLSLADYFWWIVAFSAMLLLVLSAVLARYVILLIK				
		10	20	30	40	50	60
55	orf64ng-1.pep	DRNRGVFSGQIAKRLSGMFTLVAVLPGLFLFGISQAQFNGTINSWFGNDTHEALERSNL					
	orf64-1	DRRDGVFSGQIAKRLSGMFTLVAVLPGLVFLGVSAQFNGTINSWFGNDTHEALERSNL					
		70	80	90	100	110	120
60	orf64ng-1.pep	SKSALDLAADNAVSNAPVQIDLIGTASLSGNGMSVLEHYAGSGFAQLALYNAASGKIEK					
	orf64-1	SKSALNLAADNALNAVFPQIDLIGASLSEGMSGRVLEHYAGSGFAQLALYNAASGKIEK					
		130	140	150	160	170	180
65	orf64ng-1.pep	SKSALNLAADNALNAVFPQIDLIGASLSEGMSGRVLEHYAGSGFAQLALYNAASGKIEK					
	orf64-1	SKSALNLAADNALNAVFPQIDLIGASLSEGMSGRVLEHYAGSGFAQLALYNAASGKIEK					
		130	140	150	160	170	180

		190	200	210	220	230	240
	orf64ng-1.pep	SINPHQFDQFLPDKHEWEGIQQTGVSRSIESIGGVLYAQGWLSAGTHNGRDYALFFRQPI					
5	orf64-1	SINPHKLDQPPFGKARKEIKRAGSVRDLESIGGVLYAQGWLSAGTHNGRDYALFFRQPV					
		190	200	210	220	230	240
	orf64ng-1.pep	PENVAQDAVILIEKARAKYAELSYSKKGLOTFFLVLTLLIASLLSIFLALVMALYFARRVE					
10	orf64-1	PKGVAEDAVILIEKARAKYAELSYSKKGLOTFFLATLLIASLLSIFLALVMALYFARRVE					
		250	260	270	280	290	300
	orf64ng-1.pep	PILSLAEGAKAVAQGDFSQTRPVLNDEFGRLLTKLFNHMTQELSAKEADENNRREEAA					
15	orf64-1	PVLSLAEGAKAVAQGDFSQTRPVLNDEFGRLLTKLFNHMTQELSAKEADENNRREEAA					
		310	320	330	340	350	360
	orf64ng-1.pep	RHYLCVLDGLTTGGVVVFDEKGRLLTKFNKAAEQILGMPLPFWGSSRHGWGVSQAQSSL					
20	orf64-1	RHYLCVLEGLTTGGVVVFDEQGLTKFNKAAEQILGMPLPFWGSSRHGWGVSQAQSSL					
		370	380	390	400	410	420
	orf64ng-1.pep	AEVFAAIGAAAGTDKPFVQVEYAAPDDAKILLGKATVLPEDNGNGVVMVDDITVLIRAQK					
25	orf64-1	AEVFAAIGAAAGTDKPFVHVYAAPDDAKILLGKATVLPEDNGNGVVMVDDITVLIRAQK					
		430	440	450	460	470	480
	orf64ng-1.pep	EAAMGEVAKRLAHEIRNPLTPIQLSABRLAWKLGGLDDQDAQILTRSTDITIKOVAALK					
30	orf64-1	EAAMGEVAKRLAHEIRNPLTPIQLSABRLAWKLGGLDDQDAQILTRSTDITIKOVAALK					
		490	500	510	520	530	540
	orf64ng-1.pep	EMVEAFRNYARAPSLKLENQDNLALIGDVLALYEAAGPCRFAELAGEPLTVAADTTAMRQ					
35	orf64-1	EMVEAFRNYARAPSLKLENQDNLALIGDVLALYEAAGPCRFAELAGEPLTVAADTTAMRQ					
		550	560	570	580	590	600
	orf64ng-1.pep	VLHNIFKNAAEAAEADMPFVVKSETGQDGRIVLTVCDNGKGFGRMLHNAFEPVYTDK					
40	orf64-1	VLHNIFKNAAEAAEADVPFVVKSETGQDGRIVLTVCDNGKGFGRMLHNAFEPVYTDK					
		610	620	630	640	650	660
	orf64ng-1.pep	PAGTGLGLPVVKKIIIEHGGRISLSNQDAGGACVRIILPKTVETVYAX					
45	orf64-1	PAGTGLGLPVVKKIIIEHGGRISLSNQDAGGACVRIILPKTVETVYAX					
		670	680	690	700		
	orf64ng-1.pep	ISALATFLILMGLTTPVVPVTHQVVIS----					
50	orf64-1	ISALATFLILMGLTTPVVPVTHQVVIS----					

Furthermore, ORF64ng-1 shows significant homology to a protein from *A. caulinodans*:

55	sp Q04850 NTRY_AZOCA NITROGEN REGULATION PROTEIN NTRY >gi 77479 pir S18624 ntry protein - Azorhizobium caulinodans >gi 38737 (X63841) NtrY gene product [Azorhizobium caulinodans] Length = 771 Score = 218 bits (550), Expect = 7e-56 Identities = 195/720 (27%), Positives = 320/720 (44%), Gaps = 58/720 (8%)
60	Query: 7 IAAICAVVILYGLTAATGSTSSLDYFWWIXXXXXXXXXXXXXXRYVILLKDRNRG 66 I+A+ ++L GLT + + R + K R G Sbjct: 35 ISALATFLILMGLTTPVVPVTHQVVIS----VLLNAAAVLILSAMVGREIWRIAKARARG 90
65	Query: 67 FGSQIAKRISGMFTLVAVLPGLFLGISAQINGTINSWFGNDHEALERSNLNLSKSLD 126 +++ R+ G+F +V+P + + +++ ++ ++ WF T E + S+++++ + Sbjct: 91 AAARLHIRIVGLFAVVSVPAILVAVASLTLDRLDRWFSMRTQEVIVASSVSAQTVYR 150

Query: 127 LAADNAVSNVFPVQIDLIGTASLSGNMGSVLEHYAG--SGFAQLALVNAASGKIEKSINP 184
 A N + + + DL S+ + + Y G S F Q+ AA + + +
 Sbjct: 151 EHAINIRGDIAMSADLTRLSKV-----YEGDRSRFNQILTAQAALRNLPGLMI 200

5 Query: 185 HQFDQLPDKHEHWQIQQTGSVRSLESIGGVLYAQGWLSAGTHNGRDYA-----233
 + D + + + + I + V + + IG Q + N DY
 Sbjct: 201 RR-DLSVVERAN-VNIGREFIVPANLAIGDATPDQVYILP--NDADYVAADVFLKDYDD 256

10 Query: 234 --LFFRQPIPENVAQDAVLEIKARAKYAELSYSKKGQLQTFVLVTXXXXXXXKXVMA 291
 L+ + I V ++ A Y L + G+Q F +
 Sbjct: 257 LYLVARLDPRVIGYLKTTQETLADYRSLEERRFGVQVAFALMYAVITLIVLSAVWL 316

15 Query: 292 LYFARRFVEPILSLAEGAKAVAQGDGFSQTRPVLRLND--EFGRLTKLFNHTMQLSIXXXXX 350
 L F+ + V PI L A VA+G+ P+ R + + L + FN MT +L
 Sbjct: 317 LNFSKWLVAPIIRRLMSAADHVAEGNLDRVPIYRAEGLDLSAETFNKMTHELRSQREAI 376

20 Query: 351 XXXXXXXXXXXXHYLECVLDTTGVVVFDEKGRKLTFNKAAEQILGMPLAPLNGSSRRHWG 410
 + E VL G+ GV+ D + R+ N++AE+LG L+ + RH
 Sbjct: 377 LTARDQIDSRRRFTEAVLSGVGAGVIGLDSQERITILNRSABRLG--LSEVEALHRLA 434

25 Query: 411 HGVSAQSQLLAEVFXXXXXXTDKFVQVEYAAPDDAKILLGKATVLPEDNG--NGVVM 467
 V LL E + VO D + + V E + +G V+
 Sbjct: 435 EVVPETAGLLEA-----EHARQSRVQGNITLTRDGRERVFAVRVTTEQSFEEHGVV 488

30 Query: 468 VIDDITVLIRAQKEAWGEVAKRIAEIRNPLTFIQLSAERLAWLGGKLDQDQAIQLTR 527
 +DDIT LI AQ+ +AW +VA+R+AEI+NPLTFIQLSAERL K G + QD +I +
 Sbjct: 489 TLDITELISAQRTSAWADVARRIAEIKNPLTFIQLSAERLKRKFRGRV-TQDREIFDQ 547

35 Query: 528 STDITIKQVAALKEMVEAFNRYPARAFSLKLENQDNLALIGDVLALYEAGPCRFAELAGE 587
 TDITII+QV + MV+ F +AR P + + +QD+ +I + L G +
 Sbjct: 548 CTDITIRQVGDIGRMVDEFFSSFARMKFPVVDSDQMSSEIIRQTVFLMRVGHPEVVPDSEVP 607

40 Query: 588 PLMMAA-DTTAMRQVLNHIKFNXXXXXXDMPEVRVK-----SETGDGRIVLTCD 639
 P M A D + Q L NI KN P+VR + + G+D +V+ +D
 Sbjct: 608 PAMPARFDRRLVSQALTNILNAAEAIEAVP-PDVRGQGRIRVSANRVGD--LVIDIID 664

Query: 640 NGKGFGKEMLHNAFEPYVTDKPAFTGLGLPVVKKIIEHGGGRISLSNOGAG-GACVRIIL 698
 NG G +E + EPYVT + GTGLGL +V KI+ EHGG I L++ G GA +R+ L
 Sbjct: 665 NGTGLPQESRNLLEPYVTTREKGTGLGLAIVGKIMEHGGGGLIENDAPEGRGAWIRLTL 724

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 31

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 259>:

```

1 ATGTACGCAT TTACGCGCGC ACAGCAACAG AAGGCATCTT TCCGGCTGGT
51 GCTTTTTCAT ATCCTCATCA TCGCGCCGAG CAACATCTGT TGCAGTCTCC
101 CTTTCACAAAT TTTCCGCAATC CACACCACTT GGGCGCATTT TCTCTTCC
151 TTTCACTCTCC TTGCGACCGA CCGTACCGCTC CGCATTTTGC GTTCTCACTT
201 GGCACGGCGG ATTATCTTTT GGGTGATGTT CCGCGCCTT TTGCTTTCCT
251 ACGTCTTTTC CGTTTGTGTC CACAACGGCA GTTGGACAGG CTGGGGGGGG
301 CTGTCCGAAT TCAACACCTT TGTCCGACGC ATCCGCTTAG CCAGCTTTGC
351 CGCTACGCGC ATCGGACAAA TCCTTGATAT TTTTGATTC AACAAATTAC
401 GCGCTCTGAA ACGCTGGTGG ATTGCACCGA ACGCATCAAC CGTCACTGGG
451 CACGCTTGG ATACG...

```

This corresponds to the amino acid sequence <SEQ ID 260; ORF66>:

```
1 MYAFTAAQQQ KALFRVLVLF ILIIAASNYL VQFFFIQFI HTTWGAFSFP
```

51 FIFLATDLTV RIFGSHLARR IIFWVMFPAL LLSYVFSVL F HNGSWTGLGA
101 LSEFNTFVGR IALASFAAYA IGQILDIFV F NKLRRLKAWW IAPNASTVIG
151 HALDT...

Further work revealed the complete nucleotide sequence <SEQ ID 261>:

5 1 ATGTACGCAT TTACCGCGCG ACAGCAACAG AAGGCACTCT TCCGGCTGGT
51 GCTTTTTCAT ATCCTCATCA TCGCCGCCAG CAACATATCTG GTGCAGTTCC
101 CTTTCCAAT TTTCCGCATC CACACCACTT GGGCGCATTT TTCCTTTCCC
151 TTCTCATCTC TTGCCACCGA CCTGACCGTC CGCATTTTGG GTTCTCATCT
10 201 GGCACGGCGG ATTATCTTTT GGGTGATGTT CCCCGCCCTT TTGCTTTTCT
251 ACGTCTTTTC CGTTTGTGTC CACAACGGCA GTTGACACAGG CTTGGGCGCG
301 CTGTCCGAAT TCAACACCTT TGTGCGACGC ATCGCCTTAG CCAGCTTTGC
351 CGCCTACGCG ATCGGACAAA TCCTTGATAT TTTTGTATTC AACAAATTAC
401 GCGCTCTGAA AGCGTGGTGG ATTGACACGA CCGCATCAAC CGTCATCGCG
15 451 AACCGCTTGG ATACGCTGGT ATTTTTCGCC GTTGCTTCTC ACGCAACGAC
501 CGATGGATT ATGGCGGCAA ACTGGCAGGG CATCGCTTTT TCGATTACCC
551 TGTTCAAACT TACCGTCTGC ACCTCTCTCT TCCTGCGCGC CTACGCGGTG
601 ATACTGAATC TGCTGACGAA ARAACTGACA ACCCTGCAAA CCAACACGGC
651 GCAAGACCGC CCGGCGCCCT CGCTGCAAAA TCGCTAA

This corresponds to the amino acid sequence <SEQ ID 262; ORF66-1>:

20 1 MYAFTAACQO KALFRLVLPH ILIIAASNYL VQFFQIFGI HTTWGAFSFP
51 FIFLATDLTV RIFGSHLARR IIFWVMFPAL LLSYVFSVL F HNGSWTGLGA
101 LSEFNTFVGR IALASFAAYA IGQILDIFV F NKLRRLKAWW IAPTASTVIG
151 NALDTLVEFA VAFYASSDGF MAANWQGIAP VDYLKLTVC TLFFLPAYGW
201 ILNLLTKKLT TLQTKQAQDR PAPSQNE*

25 Computer analysis of this amino acid sequence gave the following results:

Homology with the hypothetical protein o221 of *E. coli* (accession number P37619)

ORF66 and o221 protein show 67% aa identity in 155aa overlap:

30 orf66 1 MYAFTAACQOKALFRLVLPHILIIAASNYLVQFFQIFGIHTTWGAFSFPFIFLATDLTV 60
M F+ Q+ KALF L LPH+L+I +SNLVQ P I G HTTWGAFSFPFIFLATDLTV
o221 1 MNVFSQTRYKALFWLSFLHLVITSSNYLVQLPVSILGSHHTWGAFSFPFIFLATDLTV 60
orf66 61 RIFGSHLARRIIFWVMFPALLSYVFSVLPHNGSWTGLGALSEFNTFVGRIALASFAAYA 120
RIFG+ LARRII F VM PALL+SYV S LF+ GSW G GAL+ FN EV RIA ASF AYA
o221 61 RIFGAPLARRIIFAVMIPALLISYVISSLFYMGSGQGGALAHFNLTAVARIATASFAYA 120
35 orf66 121 IGQILDIFVFNKLRRLKAWWIAPNASTVIGHALDT 155
+GQILD+ VFN+LR+ + WW+AP AST+ G+ DT
o221 121 LGQILDVHVFNRLQRSRWLAPTASTLEGNVSDT 155

40 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF66 shows 96.1% identity with a 155aa overlap with an ORF (ORF66a) from strain A of *N. meningitidis*:

45 orf66.pep MYAFTAACQOKALFRLVLPHILIIAASNYLVQFFQIFGIHTTWGAFSFPFIFLATDLTV
orf66a MYAFTAACQOKALFWLVLPHILIIAASNYLVQFFQIFGIHTTWGAFSFPFIFLATDLTV
10 20 30 40 50 60
50 orf66.pep RIFGSHLARRIIFWVMFPALLSYVFSVLPHNGSWTGLGALSEFNTFVGRIALASFAAYA
orf66a RTFGSHLARRIIFWVMFPALLSYVFSVLPHNGSWTGLGALSEFNTFVGRIALASFAAYA
70 80 90 100 110 120
55 orf66.pep IGQILDIFVFNKLRRLKAWWIAPNASTVIGHALDT
:|||||:|||||:|||||:|||||:|||||:|||||

orf66a LGQILDIFVFNKLRLKAWWVAPTASTVIGNALDTLVFFAVAFYASSDGFMAANWQGIAF
130 140 150 160 170 180

orf66a VDYLEFKLTVCGLFFLPAYGVILNLLTKKLTTLQTKAQDRPAPSLQNPX
190 200 210 220

The complete length ORF66a nucleotide sequence <SEQ ID 263> is:

```

1  ATGTACGCAT  TTACCGCGCG  ACAGCAACAG  AAGGCATCT  TCTGGCTGGT
51  GCTTTTTCAT  ATCCTCATCA  TCGCGCGCAG  CAACTATCTG  GTGCAGTTCC
101  COTTCCAAAT  TTCCGCGCAT  CACACCACTT  GGGCGCGGTT  TTCCTTTCCC
151  TTCATCTTCC  TCGCCACCGA  CTTGACCGCT  CGCATTTTCC  GTTCGCACCT
201  GGCACGCGCG  ATTATCTTTT  GGGTCATGTT  CCGCGCCCTT  TTGCTTTTCC
251  ACGTCTTTTC  CGTTTGTGTC  CACAACGGCA  GTTGGACGGG  CTTGGCGCGG
301  CTGTCGGAAT  TCAACACCTT  TGTGCGGACG  ATCGCGCTGG  CAAAGTTTGG
351  CGCTACGCG  CTCGGACAAA  TCCTTGATAT  TTTTGTGTC  AACAAATTAC
401  GCGCTCTGAA  AGCGTGGTGG  GTTGCCCGCA  CTGATCAAC  CGTCACTGGG
451  AACGCCTTAG  ATACGTGGT  ATTTTTCGCC  GTTGCCTTCT  ACGCAAGCAG
501  CGATGGATTT  ATGCGCGCAA  ACTGGCAGGG  CATCGCTTTT  GTCGATTACC
551  TGTTCAAACT  CACCGTCTGC  GGTCTGTTT  TCTGCGCCGC  CTACGGCGTG
601  ATTCTGAATC  TGTCTGAGAA  AAACTGAGG  ACCCTGCAAA  CCAAACAGGC
651  GCAAGACGCG  CCGCGCGCCT  CGCTGCAAAA  TCCGTAA

```

This encodes a protein having amino acid sequence <SEQ ID 264>:

```

1  MYAFTAAQQQ  KALEFWLVLFH  ILIIAASNYL  VQFFQISGI  HTTWGAFSFP
51  FIFLATDLTV  RIFGSHLARR  IIFWVMPAL  LLSYVESVLF  HNGSWTGLGA
101  LSEFNTFVGR  IALASFAAYA  LGQILDIFV  NKLRLKAWW  VAPTASTVIG
151  NALDTLVFFA  VAFYASSDGF  MAANWQGI  VDYLEFKLTV  GLFFLPAYGV
201  ILNLLTKKLT  TLQTKAQDR  PAPSLQNP*

```

ORF66a and ORF66-1 show 97.8% identity in 228 aa overlap:

```

10 20 30 40 50 60
orf66a.pep MYAFTAAQQQKALEFWLVLFHILIIAASNYLVQFFQISGIHTTWGAFSFPFIFLATDLTV
orf66-1 MYAFTAAQQQKALFRVLVLFHILIIAASNYLVQFFQISGIHTTWGAFSFPFIFLATDLTV
10 20 30 40 50 60
70 80 90 100 110 120
orf66a.pep RIFGSHLARRIIFWVMPALLLSYVESVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA
orf66-1 RIFGSHLARRIIFWVMPALLLSYVESVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA
70 80 90 100 110 120
130 140 150 160 170 180
orf66a.pep LGQILDIFVFNKLRLKAWWVAPTASTVIGNALDTLVFFAVAFYASSDGFMAANWQGI
orf66-1 LGQILDIFVFNKLRLKAWWIAPTASTVIGNALDTLVFFAVAFYASSDGFMAANWQGI
130 140 150 160 170 180
190 200 210 220 229
orf66a.pep VDYLEFKLTVCGLFFLPAYGVILNLLTKKLTTLQTKAQDRPAPSLQNPX
orf66-1 VDYLEFKLTVCGLFFLPAYGVILNLLTKKLTTLQTKAQDRPAPSLQNPX
190 200 210 220

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF66 shows 94.2% identity over a 155aa overlap with a predicted ORF (ORF66.ng) from *N. gonorrhoeae*:

```

orf66.pep MYAFTAAQQQKALEFWLVLFHILIIAASNYLVQFFQISGIHTTWGAFSFPFIFLATDLTV 60
orf66ng MYAFTAAQQQKALEFWLVLFHILIIAASNYLVQFFQISGIHTTWGAFSFPFIFLATDLTV 60
orf66.pep RIFGSHLARRIIFWVMPALLLSYVESVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA 120

```

```

orf66ng      RIFGSHLARRIIFWVMFPALSLSYVFSVLFHNGSWTGLGAPSQFNTFVGRIALASFAAYA 120
orf66.pep    IGQILDIFVFNKLRRLKAWWIAPASTVIGNALDTLVFFAVAFYASSDEFMAANWQGIAP 155
5 orf66ng    LGQILDIFVFDKLRRLKAWWIAPASTVIGNALDTLVFFAVAFYASSDEFMAANWQGIAP 180

```

The complete length ORF66ng nucleotide sequence <SEQ ID 265> is:

```

1  ATGTACGCAT  TGACGCGCGC  ACAGCAACAG  AAGGCACCTCT  TCCGGCTGGT
51  GCTTTTCCAT  ATCCTCATCA  TCGCGCGCAG  CAACATATCTG  GTGCAGTTTCC
101  CCTTCGGGAT  TTTCCGCATC  CACACCACTT  GGGCGCGGTT  TTCCITTCOC
151  TTTCACTTCC  TCGCCACCGA  CCTGACCGTC  CGCATTTTCG  GTTCGCACCTT
201  GCGCGGCGG  ATTATCTTTT  GGGTGATGTT  CCCCGCCCTT  ttgCTTTeat
251  aCGTCTTTTC  CGTTTTGTTC  CACAACGCGA  GTTGAACGGG  CTTGGGCGCG
301  ctgTCCCAAT  TCAACACCTT  TGTGGGACGC  ATCGCGCTGG  CAAGTTTTCG
351  CGCCTACGCG  CTCGGACAAA  TCCTTGATAT  TTTCTGATTCT  GACAAATTAC
401  GCGCTCTGAA  AGCGTGGTGG  ATTGCCCGCG  CCGCATCAAC  CGTCATCGCG
451  AATGCATCGG  ACACGTTAGT  ATTTTGTGCC  GTTGCCTTTT  ACGCAAGCAG
501  CGATGAATTT  ATGGCGGCAA  ACTGGCAGGG  CATCGCTTTT  GTCGATTACC
551  TGTTCAAACT  TACCGTCTGC  ACCCTCTTCT  TCTCTGCCGC  CTACGCGCGT
601  ATACTGAATC  TGCTGACGAA  AAACTGACGC  GCCTCGCAAA  CCAACAGCGG
651  GCAAGACCGC  CCGCTGCCCT  CGCTGCAAAA  TCGTAA

```

This encodes a protein having amino acid sequence <SEQ ID 266>:

```

1  MYALTAQQQ  KALFRLVLPH  ILIIAASNYL  VQFFFRIFGI  HTTWGAFSFP
51  FIFLATDLTV  RIFGSHLARR  IIFWVMFPAL  LSYVFSVLF  HNGSWTGLGA
101  PSQFNTFVGR  IALASFAAYA  LGQILDIFV  DKLRRLKAW  IAPAASTVIG
151  NALDTLVFFA  VAFYASSDEF  MAANWQGIAP  VDYLFLKTV  TLFFLPAYGV
201  ILNLLTKKLT  ALQTKQAQDR  PVPSLQNP*

```

An alternative annotated sequence is:

```

1  MYALTAQQQ  KALFRLVLPH  ILIIAASNYL  VQFFFRIFGI  HTTWGAFSFP
51  FIFLATDLTV  RIFGSHLARR  IIFWVMFPAL  LSYVFSVLF  HNGSWTGLGA
101  LSQFNTFVGR  IALASFAAYA  LGQILDIFV  DKLRRLKAW  IAPAASTVIG
151  NALDTLVFFA  VAFYASSDEF  MAANWQGIAP  VDYLFLKTV  TLFFLPAYGV
201  ILNLLTKKLT  ALQTKQAQDR  PVPSLQNP*

```

ORF66ng and ORF66-1 show 96.1% identity in 228 aa overlap:

```

35 orf66-1.pep  MYALTAQQQKALFRLVLPHILIIAASNYLVQFFFRIFGIHTTWGAFSFPFIFLATDLTV 60
orf66ng      MYALTAQQQKALFRLVLPHILIIAASNYLVQFFFRIFGIHTTWGAFSFPFIFLATDLTV 60
orf66-1.pep  RIFGSHLARRIIFWVMFPALLSYVFSVLFHNGSWTGLGALSFFNTFVGRIALASFAAYA 120
40 orf66ng    RIFGSHLARRIIFWVMFPALLSYVFSVLFHNGSWTGLGALSQFNTFVGRIALASFAAYA 120
orf66-1.pep  IGQILDIFVFNKLRRLKAWWIAPASTVIGNALDTLVFFAVAFYASSDGFMAANWQGIAP 180
orf66ng    LGQILDIFVFDKLRRLKAWWIAPASTVIGNALDTLVFFAVAFYASSDEFMAANWQGIAP 180
45 orf66-1.pep  VDYLFLKTVCTLFFLPAYGVILNLLTKKLTALQTKQAQDRPVPSLQNPX 229
orf66ng    VDYLFLKTVCTLFFLPAYGVILNLLTKKLTALQTKQAQDRPVPSLQNPX 229

```

Furthermore, ORF66ng shows significant homology with an *E.coli* ORF:

```

50 sp|P37619|YHHQ_ECOLI HYPOTHETICAL 25.3 KD PROTEIN IN FTSY-NIKA INTERGENIC
   REGION (Q221)
>gi|1073495|pir|[s47690 hypothetical protein o221 - Escherichia coli >gi|466607
(U00039) No definition line found [Escherichia coli] >gi|1789982 (AE000423)
55 hypothetical 25.3 kd protein in ftsy-nika intergenic region [Escherichia coli]
   Length = 221
   Score = 273 bits (692), Expect = 5e-73
   Identities = 132/203 (65%), Positives = 155/203 (76%)

```

Query: 1 MYALTAQQQKALFRLVLPHILIIAASNYLVQFFFRIFGIHTTWGAFSFPFIFLATDLTV 60

M + Q+ KALF L LFH+L+I +SNYLVP P I G HTTWGAFSPFPFIATDLTV
 MNVFSQTRYKALFELSLFHLVITSSNYLVQLFVSLILGFHTTWGAFSPFIATDLTV 60
 Query: 61 RIFGSHLARRIIFWMFPALLSYVSVLFHNGSTGLGALSQNTFVGRIALASEAAYA 120
 RIFG+ LARRIIF VM PALL+SYV S LF+ GSW G GAL+ FN FV RIA ASE AYA
 Sbjct: 61 RIFGAPLARRIIFAVMIPALLISYVISSLFYMGSMQGFALAHFNLFVARIATASFMAA 120
 Query: 121 LGQILDIFVFDKLRRLKAWIAPAASTVIGNALDTLVFFAVAFYASSDEFMAANWQGIAF 180
 LGQILD+ VF++LR+ + WW+AP AST+ GN DTL FF +AF+ S D FMA +W IA
 10 Sbjct: 121 LGQILDVHVFNRLQSRRWLAPTASTLFGNVSDTLAFFFIAFWRSPPDFAEHWMEIAL 180
 Query: 181 VDYLFLKTCTCTLFFLPAYGVILN 203
 VDY FK+ + +FFLP YGV+LN
 15 Sbjct: 181 VDYCFKVLISIVFFLEPMYGVLLN 203

Based on this analysis, including the homology with the *E.coli* protein and the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 32

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 267>:

1 ATGGTCATAA AATATACAAA TTGGAATTTT GCGAAATGTG CGATAATTGC
 51 AATTTTGTATG ATGATTTTCGT TTGAAGCGAA TGCAAAATGCA GTAAATATAT
 101 CTGAACATGT TTCAGTTGAT ACCGGACAAG GTGCGAATAT TCATAGTTT
 25 152 GTACCTAATAA ATAGTAAAC TTATTCATCT GATTTAATAA AAACGGTAGA
 201 TTTAACACAC AyyCCTACGG GCGCAAAAGC CCGAATCAAC GCCAAATATA
 251 CGGCCAGCGT ATCCCGCGCC GCGCTATTGG CGGGGTTCGG CAACCTTGCC
 301 CGCTTAGGCG CGAAATTCAG CACAAGGGCG GTCCCTATG TCGGAACAGC
 351 CcTTTTAGCC CACGACGTAT ACGAAAcTTT CAAGAAGAC ATACAGGCAC
 30 401 GAGGCTACCA ATACGACCCC GAAACCGACA AATTTGTAAA AGGCTACGAA
 451 TATAGTAATT GCCTTTGGTA CGAAGACAAA AGACCTATTA ATAGAACCTA
 501 TGGCTGCTAC GCGCTTGAT. .

This corresponds to the amino acid sequence <SEQ ID 268; ORF72>:

1 MVIKYTNLNF AKLSIIAII MYSFEANANA VKISETVSVD TGQGAQIHKF
 35 51 VPKNSKTYSS DLIKTVDLTH XPTGAKARIN AKITASVSRA GVLAVGVKLA
 101 RLGAKEFSTRA VPVVGTTALL HDVYETFEKD IQARGYQYDF ETDKFKVGYE
 151 YSNCLWYEDK RRINRTYGCY GVD. .

Further work revealed the complete nucleotide sequence <SEQ ID 269>:

1 ATGGTCATAA AATATACAAA TTGGAATTTT GCGAAATGTG CGATAATTGC
 40 51 AATTTTGTATG ATGATTTTCGT TTGAAGCGAA TGCAAAATGCA GTAAATATAT
 101 CTGAACATGT TTCAGTTGAT ACCGGACAAG GTGCGAATAT TCATAGTTT
 151 GTACCTAATAA ATAGTAAAC TTATTCATCT GATTTAATAA AAACGGTAGA
 201 TTTAACACAC ATCCCTACGG GCGCAAAAGC CCGAATCAAC GCCAAATATA
 45 251 CGGCCAGCGT ATCCCGCGCC GCGCTATTGG CGGGGTTCGG CAACCTTGCC
 301 CGCTTAGGCG CGAAATTCAG CACAAGGGCG GTTCCTATG TCGGAACAGC
 351 CcTTTTAGCC CACGACGTAT ACGAAAcTTT CAAGAAGAC ATACAGGCAC
 401 GAGGCTACCA ATACGACCCC GAAACCGACA AATTTGTAAA AGGCTACGAA
 451 TAA

This corresponds to the amino acid sequence <SEQ ID 270; ORF72-1>:

1 MVIKYTNLNF AKLSIIAII MYSFEANANA VKISETVSVD TGQGAQIHKF
 50 51 VPKNSKTYSS DLIKTVDLTH IPTGAKARIN AKITASVSRA GVLAVGVKLA
 101 RLGAKEFSTRA VPVVGTTALL HDVYETFEKD IQARGYQYDF ETDKFKVGVG
 151 *

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF72 shows 98.0% identity over a 147aa overlap with an ORF (ORF72a) from strain A of *N. meningitidis*:

5	orf72.pep	10	20	30	40	50	60
		MVIKYTNLNF	AKLSIIAILM	MYSEANAN	AVKISETV	SVDTGQGA	KIHKFV
	orf72a	MVIKYTNLNF	AKLSIIAILM	MYSEANAN	AVKISETV	SVDTGQGA	KIHKFV
10	orf72.pep	70	80	90	100	110	120
		DLIKTVDLTH	PTGAKARIN	AKITASV	SRAGVL	AGVGLAR	LGAKFSTR
	orf72a	DLIKTVDLTH	PTGAKARIN	AKITASV	SRAGVL	AGVGLAR	LGAKFSTR
15	orf72.pep	130	140	150	160	170	
		HDVYETFKED	IQARGYQY	DPETDKF	VKGVEYS	NCLWYED	KRRINRTY
	orf72a	HDVYETFKED	IQARGYQY	DPETDKF	AKVSGX		
20		130	140	150			

The complete length ORF72a nucleotide sequence <SEQ ID 271> is:

25	1	ATGGTCATAA	AATATACAAA	TTTGAATTTT	GCGAAATTGT	CGATAATTGC
	51	AATTTTGTATG	ATGTATTCGT	TTGAAGCGAA	TGCAAAATGCA	GTAAAAATAT
	101	CTGAACACTGT	TTCAAGTTGAT	ACCGGACCAAG	GTGCGAAAT	TCATAAGTTT
	151	GTACCTAAAA	ATAGTAAAC	TTATTCATCT	GATTTAATAA	AAACGGTAGA
	201	TTTAACACAC	ATCCCTACGG	GCGCAAAAGC	CGAATCAAC	GCCAAAATAA
	251	CCGCCAGCGT	ATCCCGCGCC	GGCGTATTGG	CGGGGTCGG	CAAACTGCC
	301	CGCTTAGGCG	CGAAATTCAG	CACAAGGCG	GTTCCCTATG	TCGGAACAGC
30	351	CCTTTTAGCC	CACGACGTAT	ACGAAACTTT	CAAGAAGAC	ATACAGGCAC
	401	GAGGCTACCA	ATACGACCCC	GAAACCGACA	AATTTGCAAA	GGTCTCAGGC
	451	TAA				

This encodes a protein having amino acid sequence <SEQ ID 272>:

35	1	MVIKYTNLNF	AKLSIIAILM	MYSEANANA	VKISETVSVD	TQGA	KIHKF
	51	VFNKSKTYSS	DLIKTVDLTH	PTGAKARIN	AKITASVSR	AGVGL	KLAK
	101	RLGAKFSTRA	VPYVGTALLA	HDVYETFKED	IQARGYQYDP	ETDK	FAKVS
	151	*					

ORF72a and ORF72-1 show 100.0% identity in 150 aa overlap:

40	orf72a.pep	10	20	30	40	50	60
		MVIKYTNLNF	AKLSIIAILM	MYSEANAN	AVKISETV	SVDTGQGA	KIHKFV
	orf72-1	MVIKYTNLNF	AKLSIIAILM	MYSEANAN	AVKISETV	SVDTGQGA	KIHKFV
45	orf72a.pep	70	80	90	100	110	120
		DLIKTVDLTH	PTGAKARIN	AKITASV	SRAGVL	AGVGLAR	LGAKFSTR
	orf72-1	DLIKTVDLTH	PTGAKARIN	AKITASV	SRAGVL	AGVGLAR	LGAKFSTR
50	orf72a.pep	130	140	150			
		HDVYETFKED	IQARGYQY	DPETDKF	AKVSGX		
	orf72-1	HDVYETFKED	IQARGYQY	DPETDKF	AKVSGX		
55		130	140	150			

```

45      10      20      30      40      50      60
orf72ng-1.pe MVTKHTNINFAKLSIIAILMYSFEANANAVKISETLSVDTGQGAHVKEVPKSSNIYSS
|| :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
orf72-1       MWIKYTNINFAKLSIIAILMYSFEANANAVKISETLSVDTGQGAHVKEVPKSNKYTSS
      10      20      30      40      50      60

50      70      80      90      100     110     120
orf72ng-1.pe DLTKAVDLTHIPTGAKARINAKITASVSRAGVLSGVKLVQAGKFGTRAVPYVGTALLA
|| :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
orf72-1       DLIKTVDTLTHIPTGAKARINAKITASVSRAGVLAVGKVLARLGAKFSTRAVPYVGTALLA
      70      80      90      100     110     120

55      130     140
orf72ng-1.pe HDVYETFKEDIQARGCRVDEPETDKF
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

orf72-1 HDVYETFKEDIQARGYQYDPETDKFAKVSXG
130 140 150

Based on this analysis, including the presence of a putative leader sequence and transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 33

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 277>:

1 ATGAGATTTT TCGGTATCGG TTTTGTGGTG CTGCTGTTT TGGAGATTAT
51 GTCGATTGTG TGGGTTGCCG ATTGGCTGGG CGCGCGCTGG ACGTTGTTTT
101 TGATGGCGGC AGGTTTGGCC GCCGCGGTGC TGATGCTCAG GCAAACCGGG
151 GCTGACCGGT CTTTATTGGG CGGCGCGCGC AATGAGAAGC GCGCGGAAGG
201 TATCCGTTTA TCAGATGTTG TGGCCTATC..

15 This corresponds to the amino acid sequence <SEQ ID 278; ORF73>:

1 MRFFGIGFLV LLFLEIMSIV WVADWLGGW TLFLMAAGFA AGVIMLRQTG
51 LTGLLLAGAA MRSGGKVSIV QMLWPI..

Further work revealed the complete nucleotide sequence <SEQ ID 279>:

1 ATGAGATTTT TCGGTATCGG TTTTGTGGTG CTGCTGTTT TGGAGATTAT
51 GTCGATTGTG TGGGTTGCCG ATTGGCTGGG CGCGCGCTGG ACGTTGTTTT
101 TGATGGCGGC AGGTTTGGCC GCCGCGGTGC TGATGCTCAG GCAATCGGG
151 CTGTCCGCTC TTTTATTGGG GCGCGCGSCA ATAGAGAAGC GCGGAGGCT
201 ATCCGTTTAT CAGATCTTCT GGCTCTATCC TTATACGGTG CGCGCTGTGT
251 CTCTGATGAG TCGCGGATTC GTATCTCCGG TGTTGGCGST ATTGCTGCTG
301 CTGCGCTTTA AGGAGAGGGC AGTGTTGCGG GCAGGAGGTG CGGAAATATT
351 TTTCAACATG AACCAATCGG GCAGAAAGAA GGCCTTTCC CGCGATGAGC
401 ATATTATCGA GGGAGAATAT ACGGTTGAAG AGCCTTACGG CGGCAATCGT
451 TCCCGAAGAC CCATCGAACA CAAAAAGAC GAATAA

This corresponds to the amino acid sequence <SEQ ID 280; ORF73-1>:

1 MRFFGIGFLV LLFLEIMSIV WVADWLGGW TLFLMAAGFA AGVIMLRHTG
51 LSGLLLAGAA MRSGGRVSIV QMLWPIRTV AAVCLMSGF VSSVLAVLLL
101 LPFKGGAVALQ AGGAENFFNM NQSRKEGFS RDDDIIEGEY TVEEYPGGR
151 SRNAIEHKDD E*

Computer analysis of this amino acid sequence gave the following results:

35 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF73 shows 90.8% identity over a 76aa overlap with an ORF (ORF73a) from strain A of *N.meningitidis*:

		10	20	30	40	50	60
40	orf73.pep	MRFFGIGFLVLLFLEIMSIVWVADWLGGWTLFLMAAGFAAGVIMLRQTGLTGLLLAGAA					
	orf73a	MRFFGIGFLVLLFLEIMSIVWVADWLGGWTLFLMAATFAAGVIMLRHTGLSGLLLAGAA					
		10	20	30	40	50	60
		70					
45	orf73.pep	MRSGGKVSIVQMLWPI					
	orf73a	MRSGGRVSIVXMLXIRYTVAAVCXMSGPGFVSXAVLLXLPFKGGAVALQAGGAENFFNM					

The complete length ORF73a nucleotide sequence <SEQ ID 281> is:

```

1   ATGAGATTTT TCGGTATCGG TTTTGTGGTG CTGCTGTTTT TGGAGATTAT
5   51  GTCGATTGTG TGGGTTGCCG ATTGGTTGGG CGGCGGTTGG ACGCTGTTTC
101 TAATGGCGGC AACCTTTGCC GCCGGCTGCG TGATGCTCAG GCATACGGGG
151 151 CTGTCGGGTC TTTTATTGGC GGGCGCGGCA ATGAGAAGCG GCGGAGGGT
201 ATCCGTTTAT CANATGTTGT GGCNTATCCG TTATACCGTG GCGCGGCTGT
251 GTCNGATGAG TCCGGGATTC GTATCCTCGG TGTNAGCGGT ATTGCTGATG
301 CTNCCGTTTA AGGGAGGTGC AGTGTTGCAG SCACGAGGTG CGGAAAATT
351 TTTCAACATG AACCAATCGG GCAGAAAGA NGCCTTTTCC CGCGATGACG
10 401 ATATTATCGA GGGGGAATAT ACGGTTGAAG ANCCTTACGG CGGCANTCGT
451 TTCCGAAACG CCNTNGAACA CAAAAAGAC GAATAA

```

This encodes a protein having amino acid sequence <SEQ ID 282>:

```

1   MRFFGIGELV LFLFLEIMSV NVADWLGGGW TLFLEMAATFA AGVVMRLRHTG
15  1 LSGLLLAGAA MRSGGRVSIV XMLWXIRYTV AAVCXMSPGF VSSVXAVLLX
101 LFFKGGAVLQ AGGAENFFNM NXSGRKXGXS RDDDIIEGEY TVEXPYGGXR
151 LRNAHEKKD E*

```

ORF73a and ORF73-1 show 91.3% identity in 161 aa overlap

```

20  orf73a.pep      10      20      30      40      50      60
    MRFFGIGFLVLLFLEIMSVNVADWLGGGWTFLFMAATFAAGVVMRLRHTGLSLGALLAGAA
    orf73-1        10      20      30      40      50      60
    MRFFGIGFLVLLFLEIMSVNVADWLGGGWTFLFMAAGFAAGVLMRLRHTGLSLGALLAGAA

25  orf73a.pep      70      80      90     100     110     120
    MRSGGRVSIVYXMLWXIRYTVAAVCXMSPGFVSSVXAVLLLPFKGGAVLQAGGAENFFNM
    orf73-1        70      80      90     100     110     120
    MRSGGRVSIVYQMLWPRIYTVAAVCLMSPGFVSSVLAVALLLPFKGGAVLQAGGAENFFNM

30  orf73a.pep      130     140     150     160
    NXSGRKXGXS RDDDIIEGEYTVEXPYGGXRFNNAHEKKDEX
    orf73-1        130     140     150     160
    NQSGRKEGFS RDDDIIEGEYTVVEPYGGNRSNAHEKKDEX

35  orf73a.pep      130     140     150     160
    NXSGRKXGXS RDDDIIEGEYTVEXPYGGXRFNNAHEKKDEX
    orf73-1        130     140     150     160
    NQSGRKEGFS RDDDIIEGEYTVVEPYGGNRSNAHEKKDEX

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF73 shows 92.1% identity over a 76aa overlap with a predicted ORF (ORF73.ng) from *N.*

gonorrhoeae:

```

40  orf73.pep      MRFFGIGFLVLLFLEIMSVNVADWLGGGWTFLFMAAGFAAGVLMRLRHTGLTGLLAGAA 60
    orf73ng        MRFFGIGFLVLLFLEIMSVNVADWLGGGWTFLFMAATFAAGVLMRLRHTGLSGLLAGAA 60

    orf73.pep      MRSGGKVSIVYQMLWPI 76
    orf73ng        VKSGGKVSIVYQMLWPIRYTVAAVCLMSPGFVSSVLAVALLLPFKGGAVLQAGGAENFFNM 120

```

The complete length ORF73ng nucleotide sequence <SEQ ID 283> is:

```

1   ATGAGATTTT TCGGTATCGG TTTTGTGGTG CTGCTGTTTT TGGAAATTAT
5   51  GTCGATTGTG TGGGTTGCCG ATTGGCTGGG CGGCGGTTGG ACGTGTGTTT
101 TAATGGCGGC AACCTTTGCC GCCGGCTGCG TGATGCTCAG GCATACgggg
151 151 CTGTCGGGTC TTTTATTGGC TGGCGGCGCG GTAAAagta gtgGGAAGST
201 ATCTGTTTAT CagatgtGT GCGCTATCCG TTATACggtg gggcggtt
251 GTCTGatgag tCcggaATTC GPATNCTccg tgggtgcGgt ATTGCTGCTT
301 CTGCGcgttta aggGaggGgc agtgttgacg gcaggaaggtg cggaaaATTG
351 TTTCAACATg aaCcaatcgg gcagaaAaga gggatttttc cagcatgacg
55  401 atattatcga gggagaatat acggttgaaa aactgcagcg cggcaatcgt
451 tcccgaAAcg ccatcgaaca cgaaaAagac gaataA

```

This encodes a protein having amino acid sequence <SEQ ID 284>:

```

1 MRFFGIGFLV LLEFLEIMSV VVADWLGGGW TFLMAATFA AGVIMLRHTG
51 LSGLLLAGAA VKSSGKVSIV QMLWPIRYTV AAVCLMSPGF VSSVLAVLLL
101 LPFKGGAVLQ AGGAENFFNM NQSGRKGEFF HDDDIIEGEY TVEKPDGGRN
151 SRNAIEHKD E*

```

5 ORF73ng and ORG73-1 show 93.8% identity in 161 aa overlap

```

10 orf73-1.pep MRFFGIGFLVLLFLEIMSVVADWLGGGWTFLMAAGFAAGVIMLRHTGLSGLLAGAA
orf73ng MRFFGIGFLVLLFLEIMSVVADWLGGGWTFLMAATFAAGVIMLRHTGLSGLLAGAA
15 orf73-1.pep MRSGGRVSVYQMLWPIRYTVAAVCLMSPGFVSSVLAVALLLPFKGGAVLQAGGAENFFNM
orf73ng VKSSGKVSIVYQMLWPIRYTVAAVCLMSPGFVSSVLAVALLLPFKGGAVLQAGGAENFFNM
20 orf73-1.pep NQSGRKGEFFSRDDDIIEGEYTYEEFPGNRSRNAIEHKKDEX
orf73ng NQSGRKGEFFHDDDIIEGEYTYVEKPDGGRNRSRNAIEHKKDEX
25 orf73-1.pep 130 140 150 160
orf73ng 130 140 150 160

```

Based on this analysis, including the presence of a putative leader sequence and putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N. meningitidis* and *N. gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 34

The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 285>:

```

30 1 ATGTTTGT TTTCAGACGCG ATTCTT.ATG TTTCAGAAAC ATTTCAGAA
51 AGCCTCCGAC AGCGTCGTCG GAGGGACATT ATACGTGGTT GCCACGCCCA
101 TCGGCAATTT GCGGACATT ACCCTGCGCG CTTTGGCGGT ATTGCAAAAG
151 GCG..... GCGCA AGACACGCGC GTTACCAGCAG AGCTTTTGAAG
201 CGCGTACGCG ATTCAGGGCA AACTCGTCAG TGTGCGCGAA CACACGAGAC
251 GGCAGATGCG GGACAGATT GTCGGCTATC TTTACAGCGG CATGGTTTGTG
301 GCACAGGTTT CCGATGCGGG TACGCGCGCC GTGTGCGACC CCGCGCGGAA
351 ACTGCCCCGCG CGCGTGCCTG AGGCGCGGTT TAAAGTCGTT CCGTTCGTGG
401 CGCGACAC.GC GGTGATGCGG GCTTTGACCG TGGCGCGGTGT GGAAGATACC
451 GATTTTATTT TCAACGGTTT TGTACGCGCG AAATCGCGAG AACCGAGGAA
501 ACTGTTTGCC AAATGCGTGC GGGCGCGGTT TCCTATCCTC ATGTTTGA...
551 CCGCGCACCG CATCGGTGCA GCGCTTGCAG ATATGCGCGA ACTGTTCCCC
601 GAACGCGCAT TARTGCTGGC GCGCGAAATT ACGAAAACGT TTGAACGTT
651 CTTAAGCGCG ACGGTTGGGG AAATTCAGAC GGCATTGTCT GCCGACGGCG
701 ACCAATCGCG CGCGAGATG GTGTGTTGTC TTTATCCGCG GCAGGATGAA
751 AAACACGGAAG GCTTGTCCGA GTCCGCGCGA AACATCATGA AAATCCTCAC
45 801 AGCCGAGCTG CCGACCAAAC AGGCGCGGGA GCTTGCTGCC AAAATCACGG
851 GCGAGGGAAA GAAAGCTTTG TACGAT..

```

This corresponds to the amino acid sequence <SEQ ID 286; ORF75>:

```

50 1 MFVFQTAEXM FQKHLQKASD SVVGGTLYVV ATPIGNLADI TLRALAVLQK
51 A...AEDTR VTAQLLSAYG IQGKLVSVRE HNEQMADKI VGLYSDGMVV
101 AQVSDAGTFA VCDPFGAKLAR RVRASGFKVY PVYKAYAYMA ALSVAGVGG
151 DFYNGVFDP KSGERAKLEA KWRARAFIV MFEIPHRIGA ALADMELEFP
201 ERRLMLAREI TKTFETFLSG TVGEIQTALS ADGQSRGEM VLVLYPAQDE
251 KHEGLSESAQ NIMKILTAEEL PTKQAELAA KITGEGKKAL YD..

```

Further work revealed the complete nucleotide sequence <SEQ ID 287>:

```

1  ATGTTTCAGA AACATTGCA GAAAGCCTCC GACAGCGTCG TCGGAGGGAC
51  ATTATACGTG GTTCCGACGC CATTCCGCAA TTTCGCGGAC ATTACCTCGC
101 GCGCTTTGCG GGTATTGCAA AAGCGGACCA TCATCTGTGC CGAAGACACG
151 CCGTGTGCGC CACAGCTTTT GAGCGCGTAC GCATCTCAGC GCAAACTCGT
201 CAGTGTGCGC GAACACAACG AACCGGCAGAT GCGCGACAAG ATTGTCCGCT
251 ATCTTTTCAGA CGCATGTGTT GTGCGACAGG TTTCGATGCG GGGTAGCCCG
301 CCGGTGTGCG ACCCGGGCGC GAAACTCGCC CGCCGCGTGC GTGAGGCCGG
351 GTTTAAAGTC GTTCCCGTCG TGGCGCGAAG CGCGGTGATG GCGGCTTTGA
401 GCGTGGCCGG TGTGGAAGGA TCCGATTTT ATTTCACGGG TTTGTACCG
451 CCGAATCGG GAGAACGCG GAAACTGTGT GCCAAATGGG TCGGGCGCGC
501 GTTTCTCTATC GTCATGTGTT AAGCGCCGCA CGCATCGGT CGCACGCTTG
551 CCGATATGGC GGAATGTGTC CCCGAACGCC GATTATGCT GCGCGCGGAA
601 ATTACGAAAA CGTTTGAAAC GTTCTTAAGC GGCACGGTTG GGGAAATTCA
651 GACGCGATTG TCTGCCGACG GCAACCAATC GCGCGCGGAG ATGGTGTGGG
701 TGCTTTATCC GCGCGAGGAT GAAAAACACG AAGGCTTGTC CGAGTCCGCG
751 CAAAAATCA TGAATATCCT CACAGCCGAG CTGCCGACCA AACAGCGCGC
801 GGAGCTTGCT GCCAAATCA CCGCGGAGGG AAAGAAAGCT TTGTAGATTC
851 TGGCTCTGTC TTGAAAAAC AATATG

```

This corresponds to the amino acid sequence <SEQ ID 288; ORF75-1>:

```

20 1  MFQKHLQKAS DSVVGCTLYV VATPIGNLAD ITRLALAVLQ KADICAEDT
51  RVTAQLLSAY GIQKLVSVR EHNERQMA DK IVGVLSDGMV VAQVSDAGTP
101 AVCDPGAKLA RRVREAGFKV FVVVGASAVM AALSVA GVEG SDPYFNGFVP
151 PKSGERRKLF AKWVRAAFPI VMFETPHRIG ATLDMAELF PERRLMLARE
201 ITKTETFLS GTVGEIQTAL SADGNQSRGE MVLVLP AQD EKHEGLSESA
251 QNIMKILTA E LPTKQAELA AKITGEGKKA LYDLALSWKN K*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF75 shows 95.8% identity over a 283aa overlap with an ORF (ORF75a) from strain A of *N.*

meningitidis:

```

30 10 20 30 40 50 60
orf75.pep MFVFQTAFXMFQKHLQKASDSVVGCTLYVVATPIGNLADITRLALAVLQKAXXXAEDTR
orf75a MFQKHLQKASDSVVGCTLYVVATPIGNLADITRLALAVLQKADITCAEDTR
35 10 20 30 40 50
orf75.pep 70 80 90 100 110 120
orf75a VTAQLLSAYGIQKLVSVREHNERQMA DKIVGYLS DGMVVAQVSDAGTPAVCDPGAKLAR
40 60 70 80 90 100 110
orf75.pep 130 140 150 160 170 180
orf75a RRVREAGFKV FVVVGASAVMAALSVAGVEGSDPYFNGFVPPKSGERRKLF AKWVRAAFPI
45 120 130 140 150 160 170
orf75.pep 190 200 210 220 230 240
orf75a VMFETPHRIG AALDMAELF PERRLMLAREITKTETFLSGTVGEIQTAL SADGNQSRGEM
50 180 190 200 210 220 230
orf75a MFETPHRIGATLDMAELF PERRLMLAREITKTETFLSGTVGEIQTAL AADGNQSRGEM
55 250 260 270 280 290
orf75.pep VLVLYPAQDEKHEGLSESAQNIMKILTAELPTKQAELAAKITGEGKALYD
orf75a VLVLYPAQDEKHEGLSESAQNIMKILTAELPTKQAELAAKITGEGKALYDLALSWKNK
240 250 260 270 280 290
60 orf75a X

```

The complete length ORF75a nucleotide sequence <SEQ ID 289> is:

1 ATGTTTCAGA AACATTGCA GAAAGCTCC GACAGCGTCG TCGGAGGGAC
 51 ATTATACCTG GTTGCCACCG CCATCGGCAA TTTGGCGGAC ATTACCCCTG
 101 CCGCTTTGCG GGTATTGCAA AAGCGGACGA TCATCTGTGC CGAAGACACG
 151 CCGCTTACC GCGAGCTTTT GAGCGGCTAC GGCATTACAG GCAAACTCGT
 5 201 CAGCGTGGCG GAACACAACG AACCGCAGAT GCGGACACAG ATTGTCCGCT
 251 ATCTTTTACA CGGCATGGTT GTGGACACAG TTTCCGATGC GGGTACGCCG
 301 GCGGTGTGCG ACCCGGGCGC GAACTCGGCC CCGCGCTGCG GTGAGCTCGG
 351 GTTTAAAGTT GTCCTGTGTT TCGGCGCAAG CCGCGTGATG CGCGCTTTGA
 401 GTGTGGCTGG TGTGGCGGGA TCGGATTTT ATTTCACAGG TTTGTACCG
 10 451 CCGAAATCGG GCGAACGTAG GAAATTGTTT GCCAAATGGG TCGGGTGGC
 501 GTTTCCTGTC GTGATGTTTG AAACGCCGCA CCGCATCGGG CGGACCTTTG
 551 CCGATATGGC GGAATCTGTC CCCGAACGCC GATTATAGCT GCGCGCGGAA
 601 ATCACGAAAA CGTTTGAACG GTTCTTAAGC GGCACGTTG GGGAAATCA
 651 GACGGCATTC GCGGCGGACG GCAACCAATC CCGCGCGGAG ATGCTTTGCG
 15 701 TGCTTTATCC GGGCGAGGAT GAAAAACACG AAGCTTTGTC CGACTCTCGG
 751 CAAACATCA TGAAAATCCT CACAGCCGAG CTGCGCAACA AACAGCGGCG
 801 GGAGCTTGCC GCCAAATCA CCGGCGAGGG AAAAAAGCT TTGTACGATC
 851 TGGCACTGTC TTGAAAAAC AAATGA

This encodes a protein having amino acid sequence <SEQ ID 290>:

20 1 MFQKHLQKAS DSVVGGTLYV VATPIGNLAD ITRLALAVLQ KADII CAEDT
 51 RVT AQLLSAY GIQKLVSVR EHRERQMA DK IVGYS DGMV VAQVS DAGTP
 101 AVCDP GAKLA RRVREVGFKV VPVVGASAVM AALS VAGVAG SDFYFNGFVP
 151 PKSGERRKLF AKWVRVAFVP VMFETPHRIG ATLADMAELF PERRLMLARE
 201 ITKT FETFLS GTVGEIQ TAL AADGNQSRGE MVLVLYPAQD EKHEGLSESA
 25 251 QNIMKILTA E LPTKQA AELA AKITGEGRKA LYDLALS WKN K*

ORF75a and ORF75-1 show 98.3% identity in 291 aa overlap:

30 orf75a.pep 10 20 30 40 50 60
 MFQKHLQKASDSVVGGTLYVVATPIGNLADITRLALAVLQKADII CAEDTRVTAQLLSAY
 orf75-1 10 20 30 40 50 60
 MFQKHLQKASDSVVGGTLYVVATPIGNLADITRLALAVLQKADII CAEDTRVTAQLLSAY
 35 orf75a.pep 70 80 90 100 110 120
 GIQKLVSVREHNERQMA DKIVGYS DGMVVAQVS DAGTP PAVCDP GAKLARRVREVGFKV
 orf75-1 70 80 90 100 110 120
 GIQKLVSVREHNERQMA DKIVGYS DGMVVAQVS DAGTP PAVCDP GAKLARRVREAGFKV
 40 orf75a.pep 130 140 150 160 170 180
 VPVVGASAVMAALSVAGVAGSDFYFNGFVP PKSGERRKLF AKWVRVAFVP VMFETPHRIG
 orf75-1 130 140 150 160 170 180
 VPVVGASAVMAALSVAGVEGSDFYFNGFVP PKSGERRKLF AKWVRVAFVP VMFETPHRIG
 45 orf75a.pep m 190 200 210 220 230 240
 ATLADMAELFPERRLMLAREITKT FETFLSGTVGEIQ TAL AADGNQSRGEMVLVLYPAQD
 orf75-1 190 200 210 220 230 240
 ATLADMAELFPERRLMLAREITKT FETFLSGTVGEIQ TAL AADGNQSRGEMVLVLYPAQD
 50 orf75a.pep 250 260 270 280 290
 EKHEGLSESAQNIMKILTAELPTKQA AELA AKITGEGKALYDLALS WKNKX
 orf75-1 250 260 270 280 290
 EKHEGLSESAQNIMKILTAELPTKQA AELA AKITGEGKALYDLALS WKNKX

Homology with a predicted ORF from *N.gonorrhoeae*

ORF75 shows 93.2% identity over a 292aa overlap with a predicted ORF (ORF75.ng) from *N. gonorrhoeae*:

60 orf75.pep MFVFQTAFFMFQKHLQKASDSVVGGTLYVVATPIGNLADITRLALAVLQKADII CAEDTR 56
 orf75ng MSVFQTAFFMFQKHLQKASDSVVGGTLYVVATPIGNLADITRLALAVLQKADII CAEDTR 60

5	orf75.pep	VTAQLLSAYGIQKGLVSVREHNERQMDKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLAR	116
	orf75.ng	VTAQLLSAYGIQKGLVSVREHNERQMDKIVGFLSDGLVVAQVSDAGTPAVCDPGAKLAR	120
10	orf75.pep	RVREAGFKVVPVVGASAVMAALSVAGVEGSDFYFNGFVPPKSGERRKLFAKVVRAAFPV	176
	orf75.ng	RVREAGFKVVPVVGASAVMAALSVAGVAESDFYFNGFVPPKSGERRKLFAKVVRAAFPV	180
15	orf75.pep	MFETPHRIGAALADMAELFPERRILMLAREITKTFETFLSGTVGEIQTALSDAGDQSRGEM	236
	orf75.ng	MFETPHRIGATLADMAELFPERRILMLAREITKTFETFLSGTVGEIQTALAADGNSRGEM	240
20	orf75.pep	VLVLYPAQDEKHEGLSESAQNIMKILTAELPTKQAAELAAKITGEGKKALYD	288
	orf75.ng	VLVLYPAQDEKHEGLSESAQNAMKILAAELPTKQAAELAAKITGEGKKALYDLALSWKNK	300

An ORF75ng nucleotide sequence <SEQ ID 291> was predicted to encode a protein having amino acid sequence <SEQ ID 292>:

20	1	MSVFQTAFFM	FQKHLQKASD	SVVGGTLYV	ATPIGNLADI	TLRALAVLQK
	51	ADII CAEDTR	VTAQLLSAYG	IQGRILVSVRE	HNERQMDKV	IGFLSDGLVY
25	101	AQVSDAGT	PAVCDPGAKL	RVREAGFKV	VPVVGASAVMA	ALSVAGVAES
	151	DFYFNGFV	PPKSGERRKL	FAKVVRAAFP	VFETPHRIGA	ALADMAELFP
30	201	ERRILMLARE	ITKTFETFLS	GTVEIQTALA	ADGNQSRGEM	VLVLYPAQDE
	251	KHEGLSESAQ	NAMKILAAEL	PTKQAAELAA	KITGEGKKAL	YDLALSWKNK
	301	*				

After further analysis, the following gonococcal DNA sequence <SEQ ID 293> was identified:

30	1	ATGTTTCAGA	AACACTTGCA	GAAAGCCTCC	GACAGCGCTG	TCGGAGGGAC
	51	ATTATACGTG	GTTGCCACGC	CCATCGGCCAA	TTTGGCAGAC	ATTACCTCGC
35	101	GCGCTTTGGC	GGTATTGCAG	AAGCGCGACA	TCAATTTGTGC	CGAAGACACG
	151	CGCGTTACTG	CGCAGCTTTT	GAGCGCGTAC	GGCATTCAGG	CGAGTTTGGT
40	201	CAGTGTGCGC	GAACACAACG	AGCGCGCAGT	GGCGGACAG	GTAATCGGTT
	251	TCCTTTTCAGA	CGGCCTGTGT	GTGGCGCAGG	TTTCCGATGC	GGGTACGCCG
45	301	GCGCGTGTGC	ACCGGGCGGC	GAAACTCGCC	CGCCCGGTGC	CGCAAGCAGG
	351	GTTCAAAGTC	GTTCCCGTCG	TGGGCGCAGG	CGCGTAAATG	CGCGCGTTGA
50	401	GTGTGGCCGG	TGTGGCGGAA	TCCGATTTTT	ATTTCACCGG	TTTTGTACCG
	451	CCGAATCCGG	GCGAAGCTAG	GAAATTTGTT	GCCAAATGGG	TGCGGCGCGC
55	501	ATTTCTCTGC	GTATCTTTTG	AAACGCCCGCA	CCGAATCCGG	GCAACCTCTG
	551	CCGATATGGC	GGAATTTGTC	CCCGAACGCC	GTCTGATGCT	CGCGCGCGAA
60	601	ATCACGAAAA	CGTTTGAAC	GTTCTTAAGC	GGCAGCGTTG	GGGAAATTCG
	651	GACGCGATTG	CGCGCGGACG	GCAACCAATC	CGCGCGGAG	ATTGTTTGGG
65	701	TGCTTTTATC	GGCGCAGGAT	GAAACACACG	AAGGCTTGTG	CGAGTCTCGC
	751	CAAAATGCGA	TGAAATCCTT	TGCGGCGGAG	CTGCCGACCA	AGCAGCGCGC
70	801	GGAGCTTGCC	GCCAAAGATTA	CAGGTGAGGG	CAAAAGGCTT	TTGTACGATT
	851	TGGCACTGTC	GTGGAAAAAC	AAATGA		

45 This corresponds to the amino acid sequence <SEQ ID 294; ORF75ng-1>:

50	1	MFQKHLQKAS	DSVVGGLTYV	VATPIGNLAD	ITLRALAVLQ	KADII CAEDT
	51	RVTAQLLSAY	GIQGRILVSVR	EHNERQMDK	VIGFLSDGLV	VAQVSDAGTP
55	101	AVCDPGAKLA	RRVREAGFKV	VPVVGASAVM	AALSVAGVAE	SDFYFNGFVP
	151	PKSGERRKLF	AKWVRAAFPV	VMFETPHRIG	ATLADMAELF	PERRLMLARE
60	201	ITKT FETFLS	GTVEIQTAL	AADGNQSRGE	MVLVLYPAQD	EKHEGLSESA
	251	QNAMKILAAE	LPTKQAAELA	AKITGEGKKA	LYDLALSWKN	K*

ORF75ng-1 and ORF75-1 show 96.2% identity in 291 aa overlap:

55	orf75-1.pep	MFQKHLQKASDSVVGGLTYV	VATPIGNLADITLRALAVLQKADII CAEDTRVTAQLLSAY	10	20	30	40	50	60
	orf75ng-1	MFQKHLQKASDSVVGGLTYV	VATPIGNLADITLRALAVLQKADII CAEDTRVTAQLLSAY	10	20	30	40	50	60
60	orf75-1.pep	GIQKGLVSVREHNERQMDKIVGYLSDGMVVAQVSDAGT	PAVCDPGAKLARRVREAGFKV	70	80	90	100	110	120
	orf75ng-1	GIQKGLVSVREHNERQMDKIVGYLSDGMVVAQVSDAGT	PAVCDPGAKLARRVREAGFKV	70	80	90	100	110	120

[illegible]

Furthermore, ORG75ng-1 shows significant homology to a hypothetical *E.coli* protein:

sp|P45528|YRAL_ECOLI HYPOTHETICAL 31.3 KD PROTEIN IN AGAI-MTR INTERGENIC REGION (F286)

25 >gi1606086 (U18997) CRP F286 [Escherichia coli]
>gi1789535 (AE000395) hypothetical 31.3 kd protein in agai-mtr intergenic
region [Escherichia coli] Length = 286
Score = 218 bits (550), Expect = 3e-56
Identities = 128/284 (45%), Positives = 171/284 (60%), Gaps = 4/284 (1%)

30 Query: 4 KHLQKASDSVGGTLYVVTPIGNLADITLRLAVLQKADICAEADTRVTAQLLSAYGIQ 63
K Q A + S G L V + T P T P I G N L A D I T R L V L Q + I A E D T R T L L + G I
Sbjct: 2 KQKQASDANSQ--GGLYVFPPIGNLADITQRLASVLQAVDLIAEADTRHTGLLLQHPGIN 59

35 Query: 64 GRILSVENHNERKQMAKVIPLGSLDVLVQVQSDAGTFAVCDPGAKLRVREAGFGKVVVF 123
RL ++ +HNE+Q A ++ L + G + A VSDAGTF + DPG L R REAG +VVF+
Sbjct: 60 ARIFALHDHNEQKQKAEITLLAKQEQGNLALVSDAGTFPLINDPGYHLVTRCEAGIRVVF 119

40 Query: 124 VGSAVMAALS VAGVAGSDFYFNGVFPFKSGERRKLFAKWVRAAFVVMFETPHRIGATL 183
G A + ALS AG+ F + G+P K S RR ++ +E+ HR+ +L
Sbjct: 120 PGFCAAITLASSGFLPSDRFCYEGFLPAKSGRDLKALKEAEPRTLIIFYESTHRLDLS 179

45 Query: 184 ADMAELFPERR--IMLAREITKTTFETFLSGVGEIQTLAALADGNQSRGEMVLVLYPAQDEK 242
D+ + E R ++LARE+TKT+ET VGE+ + D N +GEMVL++ +
Sbjct: 180 EDIVAVLGEISRVVLARELTKTWTETINGAFVGEILLAWKWDENNRKKGEMVLIV-EGHKAQ 238

Query: 243 HEGLESQRQAMKMLIPKQAEALAAKITGEGKKALDYDAL 286
E L A + + L AELP K+A LAA+I K G A L Y AL
Sbjct: 239 EEDLPDALRLTIALLQAEPLPKKAAALAAEIHGKKNALYKYAL 282

50

Based on this analysis, including the presence of a putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 35

55 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 295>:

```

1  ATGAACAGA AAAAAACGCG TGCCGCAGTT ATTGCTGCAA TGTTGGCAGG
51 TTTTGCAGCA GC.AAAGCAC CCGAATCGA CCCGGCTTTG .....
                                     //
651 ..... ..GAGTTGG TCAGAAACCA GTTGGAGCAG GGTTTGAGAC

```

701 AGGAAAAAGC CCGCTTGAAA ATCGATGCCC TTTTGAAGA AACGGTGTC
751 AAACCGTAA

This corresponds to the amino acid sequence <SEQ ID 296; ORF76>:

5 1 MKQKKTAAAV IAAMLGFAA XKAPEIDPAL
201 ELVRNQLQQG LRQEKARLKI DALLEENGVK
251 P*

Further work revealed the complete nucleotide sequence <SEQ ID 297>:

1 ATGAAACAGA AAAAAACCGC TCGCGCAGTT ATTGCTGCAA TGTTGGCAGG
51 TTTTGGCGCA GCCAAAGCAC CCGAAATCGA CCCGGCTTTG GTGGATACGC
101 TGGTGGCGCA GATCATGCGAG CAGGCAGACC GGCATGCGGA GCAGTCCCAA
151 AAACCGGACG GGCAGGCAAT CCGAAACGAT GCCGTCCGTC GGCTGCAAA
201 TTTGGAAGTT TTGAAAAACA GGGCATTGAA GGAAGGTTTG GATAAGGATA
251 AGGATGTCCA AAACCGCTTT AAAATCGCGC AAGCGTCTTT TTATGCCGAG
301 GAGTACGTCC GTTTTCTGGA ACGTTCGGAA ACGTTTCCG AAGACGAGCT
351 GCACAAGTTT TACGAACAGC AAATCCGCAT GATCAAAATG CAGCAGGTCA
401 GCTTCGCAAC CGAAGAGGAG GCGCGTCAAG CGCAGACGT CCGTCTCAA
451 GGGCTGTCTT TTGAAGGGCT GATGAAGGCT TATCCGAACG ACGAGCAGGC
501 TTTTGACGGT TTCAATATGG CGCAGCAGCT TCCCGAGCGC CTGGCTTCGC
551 AGTTTGGCGC GATGAATCGG GCGGACGTGA CCGCGATGCC GGTCAAAATTG
601 GCGCAACGCT ATTATCTGTT CAAACTCAGC GAGGTCCGGA AAAACCCGCA
651 CGCGCAGCCT TTCGAGTTGG TCAGAAACCA GTTGAGCAGG GTTTGAGAC
701 AGGAAAAAGC CCGCTTGAAA ATCGATGCCC TTTTGAAGA AACGGTGTC
751 AAACCGTAA

25 This corresponds to the amino acid sequence <SEQ ID 298; ORF76-1>:

1 MKQKKTAAAV IAAMLGFAA AKAPEIDPAL VDTLVAQIMQ QADRHAESQ
51 KPDGGAQRND AVRRLQTLVE LKNRALKEGL DKDKDVQNRF KIAEASFYAE
101 EYVFLERSE TVSEDELHKF YEQQIRMIKL QQVSFATEEE ARQAQQLLLK
151 GLSFEGLMKR YPNDEQAFDG FIMAQQLPEP LASQFAMNR GDVTRDPVKL
201 GERYLFLKLS EVGNPDQAP FELVRNQLQQ LRQEKARLK IDALLEENGK
251 KP*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF76 shows 96.7% identity over a 30aa overlap and 96.8% identity over a 31aa overlap with an

35 ORF (ORF76a) from strain A of *N. meningitidis*:

10 20 30
orf76.pep MKQKKTAAAVIAAMLGFAAXKAPEIDPAL
|||||
orf76a MKQKKTAAAVIAAMLGFAAAXKAPEIDPALVDTLVAQIMQADRHAESQSKPDGQAIRND
10 20 30 40 50 60
//
70 80 90
orf76.pep XELVRNQLQGLRQEKARLKIDALLEENGKPKX
|||||
45 orf76a DVTRDPVKLGERYLLFKLSEVGNPDQAPFELVRNQLQGLRQEKARLKIDALLEENGKPKX
200 210 220 230 240 250

The complete length ORF76a nucleotide sequence <SEQ ID 299> is:

1 ATGAAACAGA AAAAAACCGC TCGCGCAGTT ATTGCTGCAA TGTTGGCAGG
50 51 TTTTGGCGCA GCCAAAGCAC CCGAAATCGA CCCGGCTTTG GTGGATACGC
101 TGGTGGCGCA GATCATGCGAG CAGGCAGACC GGCATGCGGA GCAGTCCCAA
151 AAACCGGACG GGCAGGCAAT CCGAAACGAT GCCGTCCGTC GGCTGCAAA
201 TTTGGAAGTT TTGAAAAACA GGGCATTGAA GGAAGGTTTG GATAAGGATA
251 AGGATGTCCA AAACCGCTTT AAAATCGCGC AAGCGTCTTT TTATGCCGAG
301 GAGTACGTCC GTTTTCTGGA ACGTTCGGAA ACGGTTTCG AAGCGGCACT
351 CGCTCAGTTT TATGAGCGGC AAATCGGCAT GATCAAAATG CAGCAGGTCA

401 GCCTTCGCAAC CGAAGAGGAG GCGCGTCAGG CGCAGCAGCT CCTGCTCAAA
 451 GGGCTGCTCTT TTGAAGGCTT GATGAAGCGT TATCCGAAGC ACAGAGCAGCG
 501 TTTTGCAGCT TTTCTTATGG CGCAGCAGCT TCCGAGCGCG CTGGCTTCGC
 551 AGTTTCGAGC GATGAATCGG GCGCAGCTTA CCGCGCATCC GGTCAAATTG
 601 GCGGACGCT ATTATCTGTT CAACTCAGC GAGGTCGGGA AAAACCCGGA
 651 CCGCGCAGCT TTCGAGTTGG TCAGAAACCA GTTGAACAA GGTTCGAGC
 701 AGGAAAAGC CCGCTTGAA ATCGATGCCA TTTTGAAGA AACCGGTGTC
 751 AAACCGTAA

This encodes a protein having amino acid sequence <SEQ ID 300>:

1 MKQKKTAAAV IAAMLAGFAA AKAPEIDPAL VDTLVAQIMQ QADRHAEQSQ
 51 KPDGQAIRND AVRRIQTLEV LKNRALKEGL DKDKDVQNR F KIAEASFYAE
 101 EYVRFLESE TVSESALRQF YERQIRMILK QQVSFATEEE ARQAQQLLK
 151 GLSFEGLMKR YPNDEQAFDG FIMAQQLPEP LASQFAAMNR GDVTRDPVKL
 201 GERYYLFKLS EVGKNPDQAP FELVRNQLQEQ GLRQEKARLK IDALLEENGW
 251 KP*

ORF76a and ORF76-1 show 97.6% identity in 252 aa overlap:

		10	20	30	40	50	60
orf76a.pep		MKQKKTAAAVIAAMLAGFAA	AKAPEIDPALVDTLVAQIMQ	QADRHAEQSQKPDGQAIRND			
orf76-1		MKQKKTAAAVIAAMLAGFAA	AKAPEIDPALVDTLVAQIMQ	QADRHAEQSQKPDGQAIRND			
		10	20	30	40	50	60
orf76a.pep		AVRRIQTLEV LKNRALKEGL	DKDKDVQNR F KIAEASFYAE	EYVRFLESE TVSESALRQF			
orf76-1		AVRRIQTLEV LKNRALKEGL	DKDKDVQNR F KIAEASFYAE	EYVRFLESE TVSEDELKHF			
		70	80	90	100	110	120
orf76a.pep		YERQIRMILK QQVSFATEEE	ARQAQQLLK GLSFEGLMKR	YPNDEQAFDG FIMAQQLPEP			
orf76-1		YERQIRMILK QQVSFATEEE	ARQAQQLLK GLSFEGLMKR	YPNDEQAFDG FIMAQQLPEP			
		130	140	150	160	170	180
orf76a.pep		LASQFAAMNR GDVTRDPVKL	GERYYLFKLS EVGKNPDQAP	FELVRNQLQEQ GLRQEKARLK			
orf76-1		LASQFAAMNR GDVTRDPVKL	GERYYLFKLS EVGKNPDQAP	FELVRNQLQEQ GLRQEKARLK			
		190	200	210	220	230	240
orf76a.pep		IDAALLEENGWVKP					
orf76-1		IDAALLEENGWVKP					
		250					

Homology with a predicted ORF from *N.gonorrhoeae*

The aligned aa sequences of ORF76 and a predicted ORF (ORF76.ng) from *N. gonorrhoeae* of the

N- and C-termini show 96.7 % and 100% identity in 30 and 31 overlap, respectively:

orf76.pep	MKQKKTAAAVIAAMLAGFAA	KAPEIDPAL	30
orf76ng	MKQKKTAAAVIAAMLAGFAA	KAPEIDPALVDTLVAQIMQ	60
orf76.pep		ELVRNQLQEQ	251
orf76ng	VTRNPVKLGERYYLFKLGAVG	KNPDAQPFELVRNQLQEQ	251

The complete length ORF76ng nucleotide sequence <SEQ ID 301> is:

1 ATGAACAGCA AAAAGACCGC TGCCGAGTT ATTGCTGCAA TGTGGCAGG
 51 TTTTGGCGCA GCCAAGCAC CGGAATCGA CCGGCTTGT GTGGATACG
 101 TGTGGCGCA GATCATGCGC CAGCAGACC GGCATGCGGA GCAGTCCCAA

5

```

151 AGACCGGACG GGCAGGCAAT CCGAAACGAT GCCGTCGCCC GGCTGCAAAAC
201 TTGTGAAGATT TTGAAAAACA GGCATTGAA GGAAGSTTTG GATAGGATA
251 AGGATGTCCA AAACCGCTTT AAAATCGCGG AAGCGTCTTT TTATGCGGAG
301 GAGTACGTCC GTTTTCTGGA ACGTTCGGAA ACGGTTTCCG AAAGCGCACT
351 CGCTCAGTTT TATGAGCGGC AAATCCGCAT GATCAAATTG CAGCAGGTCA
401 GCTTCGCAAC CGAAGAGGAG GCGCGTCAGG CGCAGCAGCT CCTGCTCAAA
451 GGGCTGTCTT TTGAAGGGCT GATGAAGCGT TATCCGAACG ACAGACGAGC
501 GTTCGACGGT TTCATTATGG CGCAGCAGCT TCCCGAGCCG CTGSGCTTcgc
551 agtttgcCGG TATGAACCGT GGCAGCGTTA CCGCAATCC GGTCAAATTG
601 GCGCAACGCT ATTACTCTGT CAAACTCGGC GCGGTCGGGA AAAACCCCGA
651 CGCGACGCTT TCGAGTTGG TCGAAACCA GTTGGACAA GSTTTGAGGC
701 AGGAAAAAGC CCGCTTGAAA ATCGATGCC TTTTGAaga Aaacggtgtc
751 AaacCGTAA

```

10

This encodes a protein having amino acid sequence <SEQ ID 302>:

15

```

1 MKQKKTAAPV TAAMLAGFAA AKAFETIDPAL VDTLVAQIMQ QADRHAEQSQ
51 RPDGQAIRND AVRRLQTLV LKNRALKRGL DKDKDVQNR F KIAEASFYAE
101 EYVRFLESE TVSESALRQF YERQIRMIKL QQVSFATEEE ARQOQLLLK
151 GLSFEGLMKR YPNDEQAFDG FIMAQQLPEP LASQFAGMNR QDVTRNPFVKL
201 GERYLFLKLG AVGNFDAQP FELVRNQLQGL GRQEKARLK IDALLEENGW
251 KP*

```

20

ORF76ng and ORF76-1 show 96.0% identity in 252 aa overlap

25

```

              10          20          30          40          50          60
orf76-1.pep  MKQKKTAAPVTAAMLAGFAAAKAFETIDPALVDTLVAQIMQQADRHAEQSQKPDGQAIRND
              |||
orf76ng      MKQKKTAAPVTAAMLAGFAAAKAFETIDPALVDTLVAQIMQQADRHAEQSQRPDGQAIRND
              |||

```

30

```

              70          80          90          100         110         120
orf76-1.pep  AVRRLQTLV LKNRALKRGLDKDKDVQNRFKIAEASFYAEYVRFLESE TVSEDELHKF
              |||
orf76ng      AVRRLQTLV LKNRALKRGLDKDKDVQNRFKIAEASFYAEYVRFLESE TVSESALRQF
              |||

```

35

```

              130         140         150         160         170         180
orf76-1.pep  YEQQIRMIKLQQVSFATEEERQAQQLLLKGLSFEGLMKRYPNDEQAFDGFIMAQQLPEP
              |||
orf76ng      YERQIRMIKLQQVSFATEEERQAQQLLLKGLSFEGLMKRYPNDEQAFDGFIMAQQLPEP
              |||

```

40

```

              190         200         210         220         230         240
orf76-1.pep  LASQFAAMNRGDVTRDPVKLGERYYLFKLSEVGKNPDAQPFELVRNQLQGLGRQEKARLK
              |||
orf76ng      LASQFAGMNRGDVTRNPFVKLGERYYLFKLGA VGKNPDAQPFELVRNQLQGLGRQEKARLK
              |||

```

45

```

              250
orf76-1.pep  IDALLEENGVKFX
              |||
orf76ng      IDALLEENGVKFX
              |||

```

50

Furthermore, ORF76ng shows significant homology to a *B. subtilis* export protein precursor:

55

```

sp|P24327|PRSA_BACSU PROTEIN EXPORT PROTEIN PRSA PRECURSOR >gi|98227|pir||S15269
33K lipoprotein - Bacillus subtilis >gi|39782 (X57271) 33kDa lipoprotein
[Bacillus subtilis]
>gi|2226124|gnl|PID|e325181 (Y14077) 33kDa lipoprotein [Bacillus subtilis]
>gi|2633331|gnl|PID|e1182997 (Z99109) molecular chaperonin [Bacillus subtilis]
Length = 292
Score = 50.4 bits (118), Expect = 1e-05
Identities = 48/199 (24%), Positives = 82/199 (41%), Gaps = 32/199 (16%)

```

60

```

Query: 70 VLKNRALKRGLDK-----DKDVQNRFKIAEASF-----YAEYVRFLESE TVSE 114
          ++ LDK DK++ N+ K + Y ++Y++ + E +++
Sbjct: 53 VLTQLVQEKVLDKKYKVS DKEIDNKLK EYKTQLG DQYTALEKQYKDKYLKEQVKYELLTQ 112

```

```

Query: 115 SA-----LRQFYERQIRMIKLQQVSFATEEEARQAQQLLLKGLSFEGIMKRYPN 163
      A      +++++E      I+      + A++ A++ L KG FE L K Y
Sbjct: 113 KAAKDNKIVTDADIKEYWEGLKGKIRASHILVADKKTAEEVEKKLKGKEDLAKEYST 172

Query: 164 DEQAFDG----FIMAQQLFEEPLASQFAAMNRGDVTRDPVKLGERYYLKFLSEVGKNPDA 218
      D A G      F      Q+ E +      + G+V+ DPKV      Y++ K +E      D
Sbjct: 173 DSSASKGGDLGWFAKEGQMDETFSKAAFLKLTGEVS-DPVKTYQGYHIIKKEERKGYDD 231

Query: 219 QPFELVRNQLPQGLRQK 237
      EL      LEQ L A
Sbjct: 232 MKKELKSEVLEQKLNDNAA 250

```

Based on this analysis, including the presence of a putative leader sequence and a RGD motif in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF76-1 (27.8kDa) was cloned in the pET vector and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 10A shows the results of affinity purification of the His-fusion protein, Purified His-fusion protein was used to immunise mice, whose sera were used for Western blot (Figure 10B), ELISA (positive result), and FACS analysis (Figure 10C). These experiments confirm that ORF76-1 is a surface-exposed protein, and that it is a useful immunogen.

Example 36

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 303>:

```

1  ATGAAAAAAT  CTTTCCTTAC  GCTGTGTTCTG  TATTGCTCTT  TACTTACCGC
51  CAGCGAAATT  GCCTTACCCC  TTGGAATTGG  GGATTGAAC  CTTACCGGCG
101  GCAAAAAATTG  CGGAAACGTT  TGCCTGACA  TTTGTGATTG  CTGCGCTGTA
151  TCTGTTTGCG  CGTAATAAAG  TGACGCGTTT  GTTGATTGCG  GTGTTTTTTG
201  CGTTCAGCAT  TATTGCCAAC  AATGTGCATT  ACGGGGATTA  TCAAAGCTGG
251  ATGAGC....  //

1201  .....  CAAACCGTAT  TCGAGCAGCT  GCAAAAGACT  CTTGACGGCA
1251  ACTGGCTGTT  TGCCATATAC  TCCGATCATG  GCCAGTATGT  TCGCCAAAGT
1301  ATCTACAATC  AAGGCACGGT  GCAGCCCGAC  AGCTATCTCG  TGCCGCTAGT
1351  GTTGTACAGC  CCGGATAAAG  CCGTGCAACA  GCGTGCCAAC  CAGGCTTTTG
1401  CGCCTTGCGA  GATTGCGCTT  CATCAGCAGC  TTTCAACGTT  CCTGATTAC
1451  ACGTTGGGCT  ACGATATGCC  GGTTTCAGGT  TGTGCGGAAG  GCTCGGTAC
1501  GGGCAACCTG  ATTACGGGTG  ATGCAGGCAG  CTTGAACATT  CGCGACGGCA
1551  AGGCGGAATA  TGTTTATCCG  CAATGA

```

This corresponds to the amino acid sequence <SEQ ID 304; ORF81>:

```

1  MKKSFLTLVL  YSLLTASEI  AYPLELGIET  LPAAKIAETF  ALTFVIAALY
51  LFARNKVTRL  LIAVFFAFSI  IANNVHYADY  QSWMT.....

401  ...QTVFSQL  QKTPDGNWLF  AYTSDRGQYV  RQDIYNQGTV  QPDSYLVVLV
451  LYSFPAKVVQ  AANQAFAPCE  IAFHQQLSTF  LIHLTGYDMP  VSGCREGVSV
501  GNLTGDAGS  LNIRDGKAEY  VYFQ*

```

Further work revealed the complete nucleotide sequence <SEQ ID 305>:

```

1  ATGAAAAAAT  CTTTCCTTAC  GCTGTGTTCTG  TATTGCTCTT  TACTTACCGC
51  CAGCGAAATT  GCCTATCGCT  TTGTATTGCG  GATTGAAACC  TTACCGGCGG

```

101 CAAAAATTGC GGAACGTTT GCGGTGACAT TTGTGATTGC TGCCTGTAT
 151 CTGTTTCGCG GTTATAAGCT GACGCGTTTG TTGATTGCGG TGTTTTTTGC
 201 GTTCATCGATT ATTGCCAACA ATGTGCTTAT CGCGGTTTAT CAAGAGCTGGA
 251 TGACGGGCAT CAATTATTGG CTGATGCTGA AAGAGTGTCT CGAAGTCGGC
 301 AGCGCGGGTG CGTCGATGTT GGATAAGTTG TGCGTGCCTG TGTGTGGGGG
 351 CGTGTGGGAA GTCATGTTGT TTTGCAGCCT TGCCAAGTTC GCCCGTAAGA
 401 CGCATTTTTC TGCGCATATA CTGTTTGCCCT TCTAATGCT GATGATTTTC
 451 GTGCGTTCGT TCGACACGAA ACAAGAGCAC GGTATTTGCG CCAAACCCAG
 501 ATACAGCCCG ATCAAAGCCA ATTATTTTTCAG CTTCCGTTAT TTTCGTGCAG
 551 GCGTGTGCG GTATCAGTTG TTTGATTTAA GCAGGATCCA CGCCTTTAAG
 601 CAGCCTGCTC CAAGCAAAAT CGGGCAGGGC AGTGTTCAAA ATATCGTCTC
 651 GATTATGGCG GAAAGCGAAA GCGCGGCCGA ITTGAAGCTG TTTGGCTACG
 701 GACGCGAAAC TTGCGCGTTT TTAACCCGGC GTTCGCAAGC CGATTTTAAAG
 751 CCGATTGTGA ACAAAGTTA TTCCGCGAGC TTTATGACTG CAGTGTCCCT
 801 GCCCAGTTT TTTCAATGCGA TACCGCACGC CACCGGCTTG GAACAAATCA
 851 GCGCGGCGCA TACCAATATG TTCGCGCTCG CCAAGAGACA GGGCTATGAA
 901 ACGTATTTTT ACAGCGCGCA ACCCATCTGAT CAGATGCGCA TTTTGAACCT
 951 AATCGGTAA GATATGATAG GCGCGAAGC AGCTGCTGCC GTTGTTCGAC
 1001 ACGGCAACGG CGCATATG CCCGATGAGA AGCTGCTGCC GTTGTTCGAC
 1051 AATAATCAAT TGCAGCAGGG CAAGCATTTT ATCGTGTGCG ACCAACGCGG
 1101 TTCCAGCGCC CCATACGCGC CATTTGTGCA GCCTCAAGAT AAGATATTGC
 1151 CGGAAGCCGA TATTGTGGAT AAGTAGACGA ACACCATCCA CAAACCCGAC
 1201 CAAATGATTCT AAACCGTATT CGAGCAGCTG CAAAGCAGCG CTGACGCGAA
 1251 CTGCGTGTTC GCCTATACCT CGCATCATGG CAGTATGTT GCACAGATA
 1301 TCTACAATCA AGGCAGCGTG CAGCCCGACA GCTATCTCGT GCCGCTAGTG
 1351 TTGTACAGCC CGGATAAGCG CGTGCAACAG CTTGCCAACG AGGCTTTTTG
 1401 GCCTTGCAGG ATTGCTTCC ATCAGCAGCT TTCAACGTTT CTGATTCAAC
 1451 CGTGGGCTA CGATATGCCG GTTTTCAGTT GTCCGGAAGG CTCGGTAACG
 1501 GGCAACCTGA TTTCCGGTGA TGCAGCAGC TTGAACATTC GCGACGGCAA
 1551 GCGGAATAT GTTTATTCGC AATGA

This corresponds to the amino acid sequence <SEQ ID 306; ORF81-1>:

1 MKKSLFLTLVL YSSLLTASEI AYRFVFGIET LPAAKIAETF ALTEVIAALY
 51 LFARYKVTRL LIAVFVAFSI IANNVHYAVY QSWMTGYNW LMLKEVTVFG
 101 SGASMLDL WLFWLWGLE VMLFCSLAIE RRTKTHESADI LFAFLMIMF
 151 VRSPTQKH SPKRETYSR IKANYFSFGY FVGRVLPQL FDLSRIPAFK
 201 QPAPSKIQG SVQNIIVLMG ESESAAHKL FGYGRETSPF LTRLSDQAFK
 251 PIVKQSYAG FMTAVSLPSF FNAIPHANGL EQISGGDTNM FRLAKEQGYE
 301 TYFVSAQAEI EMALINLIGK KWIDHLIPT QLVGNGDNM PDEKLLPLED
 351 KINLQQKHF IVLHQRGSHA PYGALLQPD VFGGEADIVD KYDNTIHKTD
 401 QMIQTVFEQL QKQPDGNWLF AYTSDHQGYV RQDIYNGQTV QPDSYLVPLV
 451 LYSFPKAVQQ AANQAFAPCE IAFHQQLSTF LHITLGYDMP VSGCREGVSF
 501 GNLTGDAGS LNIRDGKPEY VYPQ*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF81 shows 84.7% identity over a 85aa overlap and 99.2% identity over a 121aa overlap with an ORF (ORF81a) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
orf81.pep		MKKSLFLTLVLVYSSLLTASEIAYPFLGDIETLPAAKIAETFFALTFVIAALYLFARNKVTRL					
orf81a		MKKSLFLTLVLYSSLLTASEIAYRFVFGIETLPAAKIAETFFALTFVIAALYLFARYKATRL					
		10	20	30	40	50	60
orf81.pep		70	80				
orf81a		LIAVFVAFSIIANNVHYAVYQSWMT					
		70	80	90	100	110	120
		LIAVFVAFSIIANNVHYAVYQSWITGINYIWLKTEIVGSGAGMSLDKILFPAALWGVLE					
			120	130	140		
orf81.pep			QTVFEQLQKTPDGNWLFAYTSDHQGYVRQD				
orf81a		IHPHANGLEIQSGGDIVDKYDNTIHKTDQMIQTVFEQLQKQPDGNWLFAYTSDHQGYVRQD					

		280	290	300	310	320	330
5	orf81.pep	150	160	170	180	190	200
		YINQGVTVQPD	SYLVPLVLYSPDKAVQQA	ANQAFAPCEIAFHQQLST	FTFLHITLGYDM	FPVSG	
	orf81a	YINQGVTVQPD	SYLVPLVLYSPDKAVQQA	ANQAFAPCEIAFHQQLST	FTFLHITLGYDM	FPVSG	
		340	350	360	370	380	390
10	orf81.pep	210	220	230			
		CREGSVTGNLIT	GDAGSLNIRDGKAEYV	YPQX			
	orf81a	CREGSVTGNLIT	GDAGSLNIRDGKAEYV	YPQX			
		400	410	420			

The complete length ORF81a nucleotide sequence <SEQ ID 307> is:

15	1	ATGAAAAAAT	CCCTTTTCGT	TCTCTTTCGT	TATTCGTCCC	TACTTACTGC
	51	CAGCGAAAT	GCTTATCGCT	TTGTATTCGG	AATTGAAACC	TTACCGGCTG
	101	CAAAAATGGC	AGAAACGTTT	GCGCTGACAT	TTGTGATGCG	TGCGCTGTAT
	151	CTGTTTGGCG	GTTATAAGGC	AACGCTTTG	TTGATTGCGG	TGTTTTCGCG
	201	GTCACGACAT	ATGTGCACCA	ATGTGCAITTA	CCGCGTTTAT	CAAAAGCTGGA
	251	TACACGGGCA	TAAATTATTG	CTGATGCTGA	AAGAGATTAC	CGAAGTTGGC
	301	CGCGCAGGG	CGTCGATGTT	GGATAAGTTG	TGGCTGCTGT	CGTTGTGGGG
	351	CGTGTGGAA	GTCATGTTGT	TTTGCAGCCT	TGCCAAGTTC	CGCGTAAGA
	401	CGCATTTTTC	TGCCGATATA	CTGTTTGCT	TCTTAATGCT	GATGATTTTC
	451	GTGCGTTCGT	TGCAACACGA	ACAAGAACAC	GCTATTTCGC	CCAAACCGAC
25	501	ATACAGCCGC	ATCAAGACCA	ATTATTTTCA	CTTCGGTTAT	TTTGTGGCAG
	551	GCGTGTGGCC	GATACAGTTG	TTTGATTAA	CAAGAGTTCC	TGTGTTCAAA
	601	CAGCCTGCTC	CAAGCAGAA	CGGCGAAGC	AGTATTCAAA	ATATCGTCCT
	651	GATTATGGGC	GAAAGCGAAA	GCGCGGCGCA	TTGAAATTG	TTTGGCTACG
	701	GCGCGGAAAC	TTCCGCGTTT	TTGACCCAGC	TTTCGCAAGC	CGATTTTAAG
	751	CCGATTGTGA	AACAAGATTA	TTCGCGAGC	TTTATGACGG	CAGTATTCCT
	801	GCCAGTTTTC	TTTAAAGTCA	TACCGCATGC	CAACGGCTTG	GACAAATCA
	851	GCGGCGGCGA	TATTTGGGAT	AAGTAGAGCA	ACACCATCCA	CAAAACCGAC
	901	CAATGATTC	AAACCGTAT	CGACAGCTTC	CAAAAGCAGC	CTGACGCGCA
	951	CTGGCTGTTT	GCTTATGCT	CCGATCATGG	CCATATGTT	CCCCAGGATA
	1001	TCTACATCA	AGGACGCTG	CAGCCGACAC	GCTATCTCGT	GCGCCTGGTG
	1051	TTGTACAGCC	CGGATAAGCG	CGTCAACAG	GCTGCCAACC	AGGCTTTTGC
	1101	GCCTTGCGAG	ATTGCGTTCC	ATCAGCAGCT	TTCAACGTTT	CTGATTACAC
	1151	CGTTGGGCTA	CGATATGCCG	GTTTCAGGTT	GTCGCGAAGG	CTCGGTAACG
	1201	GGCAACCTGA	TTACGGGTGA	TGCAGGCAGC	TTGAACATTC	GCGACGGCAA
40	1251	GCGGAATAT	GTTTATCTCG	AATGA		

This encodes a protein having amino acid sequence <SEQ ID 308>:

	1	MKKSFLVFLFL	YSSLLTASEI	AYRFVFGIET	LPAAKMAET	ALTEFVIAALY
	51	LFARYKATRL	LIAVFFAFSI	IANNVHYAVY	QSWITGINYW	LMLEITEVVG
	101	GAGASMLDKL	WLPAWGLVE	VMLECSLAKF	RRKTHFSADI	LEAFIMIMIF
45	151	VRSFDTKQEH	GISEPKPTYSR	IKANYFSFGY	FVGRVLPYQL	FDLSKIPVFK
	201	QAPSPRIQGG	SIQNIIVLIMG	RSESAHLKL	FGYGRETSPP	LTQLSQADFK
	251	PIVKQSYSAG	FMTAVSLPSF	FNVIPHANG	EQISGGDIUD	KYDNTIHKDT
	301	QMGTVFVEQL	QKQPDGNWLF	ATTSDHQGVY	QPDYILVPLV	
	351	LYSPDKAVQQ	AAQAFAFCE	IAFHQQLSTF	LHITLGYDM	VSGCREGSVT
50	401	GNLITGDAGS	LNIRDGKAEY	VYPQ*		

ORF81a and ORF81-1 show 77.9% identity in 524 aa overlap:

		10	20	30	40	50	60
	orf81a.pep	MKKSFLVFLYSSLLTASEI	AYRFVFGIET	LPAAKMAET	FALT	FVIAALYLFARYKATRL	
55	orf81-1	MKKSFLTLVLYSSLLTASEI	AYRFVFGIET	LPAAKIAET	FALT	FVIAALYLFARYKVTRL	
		10	20	30	40	50	60
	orf81a.pep	70	80	90	100	110	120
		LIAVFFAFSIIANNVHYAVY	QSWITGINYWL	LMLEITEV	GGAGASMLDKL	WLPAWGLVE	
60	orf81-1	LIAVFFAFSIIANNVHYAVY	QSWITGINYWL	LMLEITEV	GGAGASMLDKL	WLPAWGLVE	
		70	80	90	100	110	120

		130	140	150	160	170	180
	orf81a.pep	VMLFCSLAKFRKTHFSADILFAFLMLMIFVRSFDTKQEHGISPKPTYSRIKANYFSFGY					
5	orf81-1	VMLFCSLAKFRKTHFSADILFAFLMLMIFVRSFDTKQEHGISPKPTYSRIKANYFSFGY					
		130	140	150	160	170	180
	orf81a.pep	FVGRVLPYQLFDLSKIPVFKQPAPSRIGQCSQNIIVLIMGESESAHLKLFYGRETSPF					
10	orf81-1	FVGRVLPYQLFDLSRIPAFKQPAPSKIGQGSVQNIIVLIMGESESAHLKLFYGRETSPF					
		190	200	210	220	230	240
	orf81a.pep	LTQLSQADFPKPIVKQSYSGFMTAVSLPSEFNVIPHANGLESQISGGD-----					
15	orf81-1	LTRLAQADFPKPIVKQSYSGFMTAVSLPSEFNAIPHANGLESQISGGDTNMFRLAKEQGGY					
		250	260	270	280	290	300
20	orf81a.pep	-----					
	orf81-1	TYFYSAQAENEMAILNLIGKKWIDHLIQPTQLGYGNGDNMPDEKLLPLFDKINLQGGKHF					
		310	320	330	340	350	360
25	orf81a.pep	-----		290	300	310	320
	orf81-1	IVLHQGSHAPYGALLQPDQKVFGEADIVDKYDNTIHKTDQMIQTVEQLQKQPDGNWLF					
		370	380	390	400	410	420
30	orf81a.pep	AYTSDHGQYVRQDIYNQGTVPQDSYLVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTF					
	orf81-1	AYTSDHGQYVRQDIYNQGTVPQDSYLVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTF					
		330	340	350	360	370	380
35	orf81a.pep	LIHTLGYDMFVSGCREGSGVTGNLITGDAGSLNIRDGKAEYVYPQX					
	orf81-1	LIHTLGYDMFVSGCREGSGVTGNLITGDAGSLNIRDGKAEYVYPQX					
		390	400	410	420		
40	orf81a.pep	-----					
	orf81-1	-----					
		490	500	510	520		

Homology with a predicted ORF from *N.gonorrhoeae*

The aligned aa sequences of ORF81 and a predicted ORF (ORF81.ng) from *N. gonorrhoeae* of the

N- and C-termini show 82.4 % and 97.5% identity in 85 and 121 overlap, respectively:

45	orf81.pep	MKKSFLTLVLVYSSLLTASEIAYPLELGIETLPAAKIAETFALT FVIAALYLFARNKVTRL	60
	orf81.ng	MKKSFLTLVLYSSLLTASEIAYRFVFGIETLPAARKMAETFALT FMAIALYLFARYKASRL	60
50	orf81.pep	LIAVEFAFSIIANNVHYADYQSWMT	85
	orf81.ng	LIAVEFAFSMIANNVHYAVYQSWMTG	120
55	orf81.pep	QTVFEQLQKTPDGNWLFAYTSDHGQYVRQD	433
	orf81.ng	ALLQPDQKVFGEADIVDKYDNTIHKTDQMIQTVEQLQKQPDGNWLFAYTSDHGQYVRQD	433
60	orf81.pep	IYNQGTVPQDSYLVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTFLIHTLGYDMFVSG	493
	orf81.ng	IYNQGTVPQDSYIVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTFLIHTLGYDMFVSG	493
	orf81.pep	CREGSGVTGNLITGDAGSLNIRDGKAEYVYPQ	524
	orf81.ng	CREGSGVTGNLITGDAGSLNIRNGKAEYVYPQ	524

The complete length ORF81ng nucleotide sequence <SEQ ID 309> is:

1 ATGAAAAAAT CCTTTTCTG TCTCTTTCTG TATTCATCCC TACTTACGGC
 51 CAGCGAAATC GCCTATCGCT TTGTATTCCG AATTGAACCC TTACCGGCTG
 101 CAAAAATGGC GGAAACGTTT GCCTGACATC TTAGATTTGC TGCGCTGTAT
 131 GTCTTTGGCG GTTATAAGGC TTCCGGGCTG CTGATTTGGG TGTTTTCGCG
 201 GTTCACTATG APTCCACATA CGCGGTTATG CAAGCTGGGA
 251 TGCACGGTAT TAACATTTGG CTGATGCTGA AAGAGGTATC CGAAGTCGCG
 301 ACOCGGCGCG OCTCGATGTT GGAATAAGTT TGGCTGCCCT CTTTGTGGGG
 351 CGTGGCGGAA GTCATGTGTT TTTGCAGCCT TGCCAAAGTT CGCCGTAAAG
 401 CGCATTTTTC TGCOGATATA CTGTTTGCCT TCCTAATGCT GATGATTTTC
 451 GTGCGTTTCT TCGACACGAA ACAAGAGCAC GGATATTTCG CCAAACCGAC
 501 ATACACGCGC ATCAAAGCCA ATTAATTTCAG TCTCGGTTAT TTTGTCGGCG
 551 GCGTGTGTCC GTATCAGTTG TTTGATTAA GCAAGATCCC TGTGTTCAAA
 601 CAGCCTGTCT CAAGCAAAAT CGGGCAAGCG AGTATTCAAA ATATCGTCTC
 651 GATTATGGGC GAAAGCGAAA GCGCGGCGCA TTTGAAATTG TTTGTTTACG
 701 GCGCGGAAAC TTGCGCGTTT TTAACCCGCG TTTGCGCAAG CGATTTTAA
 751 CGGATTGTGA AACAAAGTTA TTCCGCGAGC TTTATGACGG CAGTATCCCT
 801 GCCCAGTTTC TTTAAGCTCA TACCGCACGC CAACGGCTTG GAACAAATCA
 851 GCGCGGCGGA TACCAATATG TTCCGCGCTCG CCAAGAGGCA GGGCTATGAA
 901 ACTGATTTT ACCTGCGCCA GCGTGAARAC CAAATGGCAA TTTTGAACCT
 951 AATCGGTAAG AATCGGTATG ACCATCTGAT TCACGCGGAG CAGTATGCTC
 1001 ACGCACAAGG CGACAATATG CCGATGAGA AGCTGTCTGC GTTGTTCGAC
 1051 AAAATCAATT TGCAGCAGGG CAGGCATTTT ATCGTGTGTC ACCAAGCGGG
 1101 TTCGCACGCC CCATATCGCG CATTGTTGCA GCCTCAAGAT AAAGTATTGC
 1151 GCGAAGCCGA TATTGTGGAT AAGTACGACA ACACCATCCA CAAAACCGAC
 1201 CAAATGATTC AAACCGTATT CGAGCAGCTG CAAAGCAGCG CTGACGGCAA
 1251 CTGGCTGTTT GCCTATACCT CCGATCATGG CAGTATGTG CGCCAGATA
 1301 TCTACAATCA AGGCACGGTG CAGCCGACGA GCTATTGTT GCCTCTGGTT
 1351 TTGTACAGCC CGGATAAGCG CGTGCAACAG CAGTCCAAAC AGGCCTTTTG
 1401 GCCTTGGCAG ATTGCTTCC ATCAGCAGCT TTCAAAGTTC CTGATTCAACA
 1451 CGTTGGGCTA CGATATGCG GTTTCAGGTT TCGCGAAGG CTCGGTAAACA
 1501 GGCAACCTGA TTACGGGCGA TGCAGCAGC TTGAACATTC GCAACGGCAA
 1551 GGCGGARTAT GTTTATCGCG AATAA

This encodes a protein having amino acid sequence <SEQ ID 310>:

1 MKKSLFVLEF YSSLITASEI AYRFVFGIET LPAAKMAETP ALTEWIAALY
 51 LEARYKASRL LIAVFFAFSM IANNVHYAVY QSMWTDINYW LMLKEVTEVG
 101 SAGASMLDKL WLPALWGVAE VMLFCSLAKF RKRTHFSADI LEAFMLLMIF
 151 VRSFDTKQEH GISPKPTYSR IKANYFSFGY FVGRVLPYQL FDLSKI PVFK
 201 QPAPSKIQGG SIQNIIVLMG ESESAHLKL FGYGRETSPP LTRLSQADFK
 251 PIVKQSYSAG FMTAVSLPSF FNVI PHANGL EQISGGDINM FRLAKEQGYE
 301 TYFYSQAEN QMALNLLGK KWIDHLIQPT QLQYNGDNM PDEKLLPLFD
 351 KINLQQRHF IVLHQRGSHA PYGALLQPD KVFGEADIVD KYDNTIHKTD
 401 QMIQTVEFQL QKQPDGNWLF AYTSDHGGYV RQDIYNQCTV QPDSYIVPLV
 451 LYSFDKAVQQ AANQAFAPCE IAFHQQLSTF LIHTLGYDMP VSGREGSVT
 501 GNLTGDAGS LNIRNGKAEY VYEQ*

ORF81ng and ORF81-1 show 96.4% identity in 524 aa overlap:

	10	20	30	40	50	60
orf81ng-1.pep	MKKSLFVLEF	YSSLITASEI	AYRFVFGIET	LPAAKMAETP	FALT	FMIAALYLFARYKASRL
orf81-1	MKKSLFVLVLY	YSSLITASEI	AYRFVFGIET	LPAAKIART	FALT	FVIAALYLFARYKVTSL
	10	20	30	40	50	60
orf81ng-1.pep	LIAVFFAFSM	IANNVHYAVY	QSMWTDINYW	LMLKEVTEVG	SAGASMLDKL	WLPALWGVAE
orf81-1	LIAVFFAFSII	IANNVHYAVY	QSMWTDINYW	LMLKEVTEVG	SAGASMLDKL	WLPVWGVLE
	70	80	90	100	110	120
orf81ng-1.pep	VMLFCSLAKF	RKRTHFSADI	LEAFMLLMIF	VRSFDTKQEH	GISPKPTYSR	IKANYFSFGY
orf81-1	VMLFCSLAKF	RKRTHFSADI	LEAFMLLMIF	VRSFDTKQEH	GISPKPTYSR	IKANYFSFGY
	130	140	150	160	170	180
orf81ng-1.pep	FVGRVLPYQL	FDLSKIPVFK	QPAPSKIQGG	SIQNIIVLMG	ESESAHLKL	FGYGRETSPP
orf81-1	FVGRVLPYQL	FDLSKIPVFK	QPAPSKIQGG	SIQNIIVLMG	ESESAHLKL	FGYGRETSPP
	190	200	210	220	230	240
orf81ng-1.pep	FVGRVLPYQL	FDLSKIPVFK	QPAPSKIQGG	SIQNIIVLMG	ESESAHLKL	FGYGRETSPP
orf81-1	FVGRVLPYQL	FDLSKIPVFK	QPAPSKIQGG	SIQNIIVLMG	ESESAHLKL	FGYGRETSPP

5	orf81-1	FVGRVL	PYQL	FDL	SRI	PAFK	QPA	PSK	IGQ	GSV	QNI	VLM	GES	ESA	AHL	KLF	GYG	RETS	SPF
		190	200	210	220	230	240	250	260	270	280	290	300	310	320	330	340	350	360
	orf81ng-1.pep	LTRL	SQAD	FKPI	VKQ	SYS	AGF	MTAV	SLPS	FFNV	IPH	ANG	LEQ	ISG	GD	TNMF	LRA	KEQ	QGYE
10	orf81-1	LTRL	SQAD	FKPI	VKQ	SYS	AGF	MTAV	SLPS	FFNA	IPH	ANG	LEQ	ISG	GD	TNMF	LRA	KEQ	QGYE
		250	260	270	280	290	300	310	320	330	340	350	360	370	380	390	400	410	420
	orf81ng-1.pep	TYFY	SAQ	ENQ	MAIL	NLIG	KWID	HLIQ	PTQL	GYG	NGD	NMP	DEK	LLPL	FDK	INL	QQG	RHF	
15	orf81-1	TYFY	SAQ	ENQ	MAIL	NLIG	KWID	HLIQ	PTQL	GYG	NGD	NMP	DEK	LLPL	FDK	INL	QQG	RHF	
		310	320	330	340	350	360	370	380	390	400	410	420	430	440	450	460	470	480
	orf81ng-1.pep	IVLH	QRG	SHAP	YGALL	QPD	KVFE	ADIV	DKY	DNTI	HKTD	QMI	QTV	FEQ	LRQ	PDG	PNW	LF	
20	orf81-1	IVLH	QRG	SHAP	YGALL	QPD	KVFE	ADIV	DKY	DNTI	HKTD	QMI	QTV	FEQ	LRQ	PDG	PNW	LF	
		370	380	390	400	410	420	430	440	450	460	470	480	490	500	510	520	530	540
	orf81ng-1.pep	AYTS	DHG	QYV	RQDI	YNQ	GTVP	QDSY	IVL	VL	YSP	DKAV	QQA	ANQ	AFAP	CEIA	FHQ	QL	STF
25	orf81-1	AYTS	DHG	QYV	RQDI	YNQ	GTVP	QDSY	IVL	VL	YSP	DKAV	QQA	ANQ	AFAP	CEIA	FHQ	QL	STF
		430	440	450	460	470	480	490	500	510	520	530	540	550	560	570	580	590	600
	orf81ng-1.pep	LIHT	LGDM	PMV	SGCR	GSV	TGN	LIT	GDAG	SLN	IRN	GKA	EYV	YYP	QX				
30	orf81-1	LIHT	LGDM	PMV	SGCR	GSV	TGN	LIT	GDAG	SLN	IRN	GKA	EYV	YYP	QX				
		490	500	510	520	530	540	550	560	570	580	590	600	610	620	630	640	650	660
	orf81ng-1.pep	LIHT	LGDM	PMV	SGCR	GSV	TGN	LIT	GDAG	SLN	IRN	GKA	EYV	YYP	QX				

Furthermore, ORF81ng shows significant homology to an *E.coli* OMP:

gil1256380 (U50906) outer membrane adherence protein-associated protein [E. coli] length = 547
Score = 87.4 bits (213), Expect = 2e-16
Identities = 122/468 (26%), Positives = 198/468 (42%), Gaps = 70/468 (14%)

Query: 25 VFGIETLPAAKMAETFA-LTFMIAALYLFARYKAS--RLLIIVFFAFSMIANNVHYAVYQ 81
Sbjct: 29 VFGITNLVASSGAHMVQRLFFVLTILVVRKISSLFLRLVLAAPFVL-LTAAADMSISLY- 86

Query: 82 SWMT-----GINYWLMLKEVTEVGSAGASMLDKLWLPALMGVAEVMFLCSLAKFRKRT 134
Sbjct: 87 SWCTEGTTFNDGFAISVLQSDPDEV----AKMLG-MYSPYLCFAFLSLFLAVIIRYDV 141

Query: 135 HFSADILFAFLMLMIFVRSF-----DTKQEHGISPKPTYRIKAN--YFSFGYFVG 183
Sbjct: 142 SLPTKKVTGILLIVISGSLFSACQFAYKDAKKNKNAFSPYILASREATYTPFNLNLYFAL 201

Query: 184 RVLPLYQ--LFDLSKIPVFKQPAKPSKIGQGSQNIQVLMGESESAHLKLFYGRGTSPPFL 241
Sbjct: 202 AAKEHQRLLSIANTVYFQL---SVRDTGIDTYVLVIGSVRVNDNMSLYGTRSTTPQV 257

Query: 242 TRLSQADFKPIVKQSYSAGFMTAVSL---SFFNVIPHANGLEQISGGDTNMFLRAKEQG 298
Sbjct: 258 E--AQRQKIKLFNQALISGAPYALSVPFLSLTADSVLSH----DIHNPYDNINMANQAG 310

Query: 299 YETFYSAQA---ENQMAILNLIGKKWIDHLIQPTQLGYGNGDNMPDEKLLPLFDKINLQ 355
Sbjct: 311 FQTFEYSSQSAFRQNGTAVTSI-----AMRAMETVYVRGF---DELLPHLSQALQQ 359

Query: 356 --QGRHFIVLHQRGSHAPYGALLQPDQKVFEADIVDK-YDNTIHKTDQMIQTVFEQLQK 412
Sbjct: 360 NTQQKLIKVLHLNGSHEPACSAYPQSSAVFPQDDQDACYDNTSHYTTDSLGGQVFELLK- 418

Query: 413 QPDGNWLFAYTSDHG---QYVRQDIYNQ--TVQPDYSIVPL-VLYSP 454
D Y +DHG +++++ G +Y VP+ + YSP

Sbjct: 419 --DRASVMYFADHGLERDPTKGNVYFHGGREASQQAYHVPFIWYSP 464

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 37

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 311>:

```

1  ...ACCGTGTCC TCTTCATCCC CCTCGTCTCT ACAC.GTGGC GGACACTGAC
51  CGGCTACTCT GCCCAAGGCG GCGGCAAGCG CTTTGGCGTC GAACAGAAGC
101 TCGTCGCCGC ATCGTCCCGC GCGGCGGTCA AAGAAATGGA TTTGTCCGCG
151 yTAAAGGAC GCAAAGCGCG CyTTTACGTC TCGGTTATGG GCGACCAAGS
201 TTCGGGCAAC ATAAGCGGCG GACGCTACTC TATCGACGCA CTGATACGCG
251 GCGGCTACCA CAACAACCCC GAAAGTGCCA CCCAATACAG CTACCCCGCC
301 TACGCACTA CCGCCACACC CAAATCGGAC CGGCTCTCCA GGTATACACC
351 TTTCCATCTG CTTTGTGAAG CCCCOCGCGC GCGYCTGACG AAAAACAGCG
401 GACGCAAGG CGAAGCGTCC GCGGACTGCT CGGTCAACGG CACGGGCGAC
451 TACGCAACG AAACCGTGCT CGCCAAACCCC CGGACGTTT CTTCTCTGAC
501 CAACCTCATC CAACCGTCT TCTACCTGCG CGGCTCGGAA TCGTACCGC
551 CGGATACGC CGACACCGAC GTATTCTGTA CGTCGACGT A...
```

This corresponds to the amino acid sequence <SEQ ID 312; ORF83>:

```

1  ..LLLLFIPLVL TXCGTLTGIL AHGGGKRFAV EQELVAASSR AAVKEMDLA
51  LKGRKAAXYV SVMGDQSGN ISGGYVSIDA LIRGGYHNHP ESATQYSYPA
101 YDTTATTKSD ALSSVTTSTS LLNAPAAXLT KNSGRKGRS AGLSVNMGTD
151 YRNETLLANP RDVSLTNLI QTVFYLRGIE VVPFYADTD VEVTVDDV..
```

Further work revealed the complete nucleotide sequence <SEQ ID 313>:

```

1  ATGAAAACCC TGCCTCTCT CATCCGCCCTC GTCCTCAGC CCTGCGGCAC
51  ACTGACCGGC ATACCGCGCC ACGGCGGCGG CAAACGCTTT GCCGTGCAAC
101 AAGAACTCGT CGCCGCATCG TCCCGCGCGC CGGTCAAAGA AATGGAATTG
151 TCCGCCCTAA AAGGACGCAA AGCCGCCCTT TACGTCCTCG TTATGGGCGA
201 CCAAGGTTCC GGCAACATAA GCGGCGGAGC CTACTCTATC GACGCACTGA
251 TACGCGCGGG CTACCAACAC AACCCCGAAA GTGCCACCCA ATACAGCTAC
301 CCGCCCTACG ACACTACCGC CACCAACAAA TCGGACCGCG TCTCCAGCGT
351 AACCACTTCC ACATCGCTTT TGAACGCGCC CGCCGCGCGC CTGACGAAAA
401 ACAGCGGACG CAAAGCGGAA CGCTCCGCGC GACTGTCGGT CAACGCAACG
451 GCGGACTACC GCAACGAAAC CTTGCTCGCC AACCCCGCGG ACCTTCTCTT
501 CCTGACCAAC CTCATCAAAC CCGTCTCTTA CTTGCGCGGC ATCGAAGTGC
551 TACCGCCCGA ATACGCGGAC ACGGACGTAT TCGTAACCGT CGACGTATTCT
601 GGCACCGTCC GACGCGGTAC CGAATGTCAC CTCTACAACG CCGAAACCGT
651 TAAAGCCCAA ACCAAGCTCG AATATTTCGC CGTTGACCGC GACAGCGCGA
701 AACTGCTGAT TACCCTTAAA ACCGCGCGCT ACGGATCCCA ATACCAAGAA
751 CAATAOGCCC TTTGGACCGG CCCTTACAAA CTGACGAAAA CGGTCAAAGC
801 CTCAGACCGC CTGATGCTCG ATTTCTCCGA CATTACCCCC TACGGCGACA
851 CAACCGCCCA AAACCGTCCC GACTTCAAAC AAAACAACCG TAAAAAACCC
901 GATGTCGGCA ACGAAGTCAT CCGCGCGCGC AAAGGAGGAT AA
```

This corresponds to the amino acid sequence <SEQ ID 314; ORF83-1>:

```

1  MKTLLLLIPL VLTACGTLTG IPAHGGGKR FAVEQELVAAS SRAAVKEMDL
51  SALKGRKAAL YVSVMGDQGS GNISGGYSI DALIRGGYHN NPESATQYSY
101 PAYDTTATTK SDALSSVTT SLLNAPAAA LTKNSGRKGE RSAGLSVNGT
151 GDIYRNETLLA NPRDVSFLTN LIQTIFYLRG IEVVPPEYAD TDVFTVVDVF
201 GTVRSRTELH LYNAETLKQA TKLEYFAVD RSKLLITPK TAAYESQYQE
251 QYALWTGFPY VSKTVKASDR LMVDFSDITP YGDTTAQNRN DFKQNNKKPK
301 DVGNEVIRRR KGG*
```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF83 shows 96.4% identity over a 197aa overlap with an ORF (ORF83a) from strain A of *N. meningitidis*:

5	orf83.pep	10	20	30	40	50
		10	20	30	40	50
10	orf83a	10	20	30	40	50
		10	20	30	40	50
15	orf83.pep	60	70	80	90	100
		60	70	80	90	100
15	orf83a	60	70	80	90	100
		60	70	80	90	100
20	orf83.pep	120	130	140	150	160
		120	130	140	150	160
20	orf83a	120	130	140	150	160
		120	130	140	150	160
25	orf83.pep	180	190			
		180	190			
25	orf83a	180	190			
		180	190			

The complete length ORF83a nucleotide sequence <SEQ ID 315> is:

30	1	ATGAAACCC	TGCTCCTCT	CATCCGCCCT	GTCTCAGAC	CCTGCGGCAC
	51	ACTGACCGGC	ATACCCGCCCT	ACGGCGCGCG	CAAAACGCTTT	GCCGTGCAAC
	101	AAGAAGCTCGT	CGCCGCATCG	TCCGCGCGCG	CGCTCAAGAA	AATGGACTTG
	151	TCCGCCCTGA	AAGGACGCA	AGCGCGCCTT	TACGCTCCG	TTATGGGCGA
	201	CCAAGTTTCG	GGCAACATAA	GGGCGGACG	CTACTCTATC	GACGCACTGA
	251	TACGCGCGCG	CTACCAACAC	AACCCGAAA	GTGCCACCA	ATACAGTCTAC
	301	CCGCGCTACG	ACACTACGCG	CACCCCAAA	TCCGCGCGC	CTCCAGCGT
	351	AACCACTTCC	ACATCGCTTT	TGAACGCCC	CGCGCGCGC	CTGACGAAA
	401	ACAGCGGACG	CAAGCGGAA	CGCTCGCGC	GACTGTCGT	CAAGCGGACG
	451	GGGACTACCG	GCAAGCAAA	CCGTGTCGC	AACCCCGCG	ACGTTCTCTT
	501	CCTGACCAAC	CTCATCCAAA	CCGTCTCTA	CCTGCGCGC	ATCGAAGTCG
	551	TACCGCCCGA	ATACGCGCAC	ACGACGATAT	TGTAACCGT	CGAAGTATTC
	601	GGCACCGTCC	GCAGCCGCAC	CGAAGTGCAC	CTCTACACG	CGAAGACCTT
	651	TAAAGCCCAA	ACCAAGCTCG	AATATTTCGC	CGTTACCGC	GACAGCCGGA
	701	AACGTGCTAT	TGCCCTTAAA	ACCGCGCCT	ACGAATCCCA	ATACCAAGAA
	751	CAATACGCCCT	CTGGGATGGG	ACCTTACAGC	GTCCGCAAAA	CCGTCAAGAC
	801	CTCAGACCGC	CTGATGGTGG	ATTTCCTCGA	CATCACCCTC	TACGCGCACA
	851	CAACGCGCCA	AAACCGTCCC	GACTTCAAC	AAACCAAGCG	TAAAGAACCC
	901	GATGTCGCGA	ACGAAGTCAT	CGCGCGCGC	AAAGGAGGAT	AA

This encodes a protein having amino acid sequence <SEQ ID 316>:

50	1	MKTLXLPLIPL	VLTAAGTLTG	IPAHGGGKRF	AVEQELVAAS	SRAAVKEMDL
	51	SALKGRKAAL	VVSVMGDS	GNISGGYSI	DALIRGGYHN	NPEATQYSY
	101	PAYDTTATTK	SDALSSVTTT	TSLLNAPAA	LTNNSGRKGE	RSAGLSVNGT
	151	GDYRNETLLA	NPDVDFLTN	LQTVFYLRG	IEVVPFYAD	TDVFTVDVF
	201	CTVRSRTELH	LYNAETLKAQ	TKLEYFAVDR	DSRKLIIAPK	TAAYESQYVE
	251	QYALWMPYS	VGKTVKASDR	LMVDFSDITP	YGDTTAQNRP	DFKQNNKKPK
	301	DVGNEVIRRR	KGG*			

ORF83a and ORF83-1 show 98.4% identity in 313 aa overlap:

	10	20	30	40	50	60
orf83a.pep	MKTLXLPLIPL	VLTAAGTLTG	IPAHGGGKRF	AVEQELVAAS	SRAAVKEMDL	SALKGRKAAL

[illegible]

Homology with a predicted ORF from *N.gonorrhoeae*

ORF83 shows 94.9% identity over a 197aa overlap with a predicted ORF (ORF83.ng) from *N.*

gonorrhoeae:

	orf83.pep	TLLLFLPLVLTXCGLTGLIHHGGGKRFVQEQLVAASSRAAVKEMDLSALKGRKAAX	58
40	orf83ng	MKTLTLLFLPLWLTACGLFTGLIFAHGGGKRFVQEQLVAASSRAAVKEMDLSALKGRKAAL	60
	orf83.pep	YVSMVGDDGSGSNISGGRYSDIALIRGGYHHNPNESATQYSYPAYDTTATTKSDALS SVTTS	118
45	orf83ng	YVSMVGDDGSGSNISGGRYSDIALIRGGYHHNPNDSATQYSYPAYDTTATTKSDALSGVTTS	120
	orf83.pep	TSLNLNAPAAXLTKNSGRKGERSAGLSVNGTGDYRNETLLANPRDVSFLTNLIQTVFYLRG	178
	orf83ng	TSLNLNAPAAALTKNNGKRGERSAGLSVNGTGDYRNETLLANPRDVSFLTNLIQTVFYLRG	180
50	orf83.pep	IEVVPXYADTDVFTVDV	197
	orf83ng	IEVVPPEYADTDVFTVDVFGVGRVSRTELHLYNAETLKAQTKLEYFAVDRDSRKLIIAPK	240

The complete length ORF83ng nucleotide sequence <SEQ ID 317> is:

55

1	ATGAAAACCTG	TGCTCTCTCT	CATCCCCCTC	GTACTCACTGC	CTTCGGGAGAC
51	ACTGACCCGGC	ATATACCCCGC	ACGGCGGGGG	CAACCGCTTTT	CGCTCGTGAAC
101	AGGAACTCTGT	CGCGCATATG	TCCGCGCGCG	CTCGTCAAAGA	AATGACTCTTG
151	TCGCCCTCTGA	AGGAGACGAA	ACGGCCCTTT	TACGTTCTCG	TTATGGGCGAC
201	CGCAAGGTTG	GTCAACATCA	GGGGCGGACG	CTACTCTCATC	GAGCAGCTTAC
251	TACGCGCGGG	GGGACACAAA	AACCCCGAGA	CGCCACCGCG	ATACAGCTAC
301	CCCGGCTATG	ACATCATCCG	CAACACCAAA	TCGACGCGCT	TCCTCGCGGCT
351	AACCATCTTC	ACATGCTCTT	TGAGACCCGC	CGCGCGCGCT	CTGACGAAAA
401	ACACGCGGAA	CGCTCTGCTG	CGCTCTGCTG	CGCTCTGCTG	CGCTCTGCTG
451	GGCAGCTACC	GCACGAAGAA	CTCTGCTCGC	ACACCCCGCG	ACGTTTCTCTT
501	CTTGACCAAC	CTCATCTGAA	CGSTCTTCTG	CTGCGGCGGG	AGTAAGTGGT

5 551 TACCGCCGCA ATAOGCCGAC ACCGACGTAT TCGTAACCGT CGACGTATTC
 601 GGCACCGTCC GCAGCCGTAC CGAACTGCAC CTCTACAACG CGAAACCGCT
 651 TAAAGCCCAA ACCAAGCTCG AATATTTCGC CGTCGACCGC GACAGCCGGA
 701 AACTGCTGAT TCCGCTTAA ACCGATCCCA ATACAAAGAA
 751 CAATACCGCC TCTGGATGGC ACCTTACAGC GTCCGCAAAA CCGTCAAGC
 801 CTCAGACCGC CTGATGCTGC ATTTCTCCGA CATCACCACC TACGGCACA
 851 CAACCGCCCA AAACCGTCCC GACTTCMAAC AAAACACGG TAAAAACCC
 901 GATCTCGCA ACGAAGTCAT CCGCCCGCG AAAGGAGGAT AA

This encodes a protein having amino acid sequence <SEQ ID 318>:

10 1 MKTLLLLLPL VLTACGTLTG IPAHGGGKRF AVEQELVAAS SRAAVKEMDL
 52 SALKGRKAAL YVSMGDQGS GNISGGYSI DALIRGGYHN NPD SATRYSY
 101 PAYDITATTK SDALSGVTTT TSLNAPAAA LTKNNGRKE RSAGLSVNGT
 151 GDYRNETLLA NFRDVSFLTN LIQTIFYLRG IEVVPPEYAD TDVFVTVDVF
 201 GTVRSRTELH LYN AETLKAQ TKLEYFAVDR DSRKLLIAPK TAAYESQYQE
 15 251 QYALWMGPYS VGKTVKASDR LMDVDFSDITP YGDTTAQNRPF DKQNNGNRP
 301 DVGNEVIRRR KGG*

ORF83ng and ORF83-1 show 97.1% identity in 313 aa overlap

		10	20	30	40	50	60
20	orf83-1.pep	MKTLLLLLPLVLTACGTLTGIPAHGGGKRF	AVEQELVAASRAAVKEMDLSALKGRKAAL				
	orf83ng	MKTLLLLLPLVLTACGTLTGIPAHGGGKRF	AVEQELVAASRAAVKEMDLSALKGRKAAL				
		10	20	30	40	50	60
25	orf83-1.pep	YVSMGDQGS	GNISGGYSIDALIRGGYHN	NPDSATRYSY			
	orf83ng	YVSMGDQGS	GNISGGYSIDALIRGGYHN	NPDSATRYSY			
		70	80	90	100	110	120
30	orf83-1.pep	TSLNAPAAALTKNSGRKGRSAGLSVNGT	GDYRNETLLANFRDVSFLTNLIQTIFYLRG				
	orf83ng	TSLNAPAAALTKNSGRKGRSAGLSVNGT	GDYRNETLLANFRDVSFLTNLIQTIFYLRG				
		130	140	150	160	170	180
35	orf83-1.pep	IEVVPPEYADTDVFVTVDVFGTVRSRTELH	LYNAETLKAQTKLEYFAVDR	DSRKLITPK			
	orf83ng	IEVVPPEYADTDVFVTVDVFGTVRSRTELH	LYNAETLKAQTKLEYFAVDR	DSRKLITPK			
		190	200	210	220	230	240
40	orf83-1.pep	TAAYESQYQEQYALWTPGYKVS	KTVKASDR	LMDVDFSDITPYGDTTAQNRPF	DKQNNGNRP		
	orf83ng	TAAYESQYQEQYALWTPGYKVS	KTVKASDR	LMDVDFSDITPYGDTTAQNRPF	DKQNNGNRP		
		250	260	270	280	290	300
45	orf83-1.pep	DVGNEVIRRRKGGX					
	orf83ng	DVGNEVIRRRKGGX					
		310					

Based on this analysis, including the presence of a putative ATP/GTP-binding site motif A (P-loop) in the gonococcal protein (double-underlined) and a putative prokaryotic membrane lipoprotein lipid attachment site (single-underlined), it is predicted that the proteins from *N. meningitidis* and *N. gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 38

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 319>:

```

1   ATGGCAGAGA TCTGTTTGAT AACCGGCACG CCCGGTTCAG GGAAAAACATT
5   51   AAAAAATGGTT TCCATGATGG CGAATGATGA AATGTTTAAAG CCTGATGAAA
101  101  AAGCCATACG CCGTAAAGTA TTTACGAACA TAAAGGCGCTT GAAATACCGG
151  151  CACACCTACA TAGAAACGGA CGCAAAAAGG CTGCCGAATG CGACAGATGA
201  201  CGAGCTTTCG GCGCATGATA TGTACGAATG GATAAAGAAG CCCGAAAATA
251  251  TCGGGTCTAT TGTCAATTGA GATGAAGCTC AAGAOSTATG GCGGCGACGC
10  301  TCGGCAGGTT CAAAAATCCC TGAAAATGTC CAATGGCTGA ATACGCACAG
351  351  ACATCAGGGC ATTGATATAT TTGTTTGGAC TCAAGGTCCT AAGCTTCTAG
401  401  ATCAAAATCT TAGAAGCGTT GTACGGAAC ATTACACAT CGCTTCAAC
451  451  AAGATGGGTA TCGCTACGCT TTTAGATGG AARATATGCG CGGACGATCC
15  501  CGTAAAAATG GCATCAAGCG CATCTCCAG TATCTATACA CTGGATAAAA
551  551  AAGTTTATGA CTTGTATsrr TmmCGGGAAG TTCATACCGT AAATAAGGTC
601  601  AAGCGGTCAA AGTGGTTTTA CACTCTGCCA GTAATAGTAT TGCTGATTCC
651  651  CGTGTTTGTC GGCCTGTCTC ATAAATGTT GAgCaGTTAC GGAAAAAAC
701  701  aGGAAGAACC CGCAGCACA GAATCGCGGG CAACAGAACA GCAGGCAGTA
20  751  CTTCCGATA AAACAGAAGG CGAGCCGGTA AATACGGCA ACCTTACCGC
801  801  AGATATGTTT GTTCCGACAT TGTCCGAAA AGCCGAgAGC AAGCGaGTTT
851  851  ATAACGGTGT AAGCCAGGTA AGAACCTTG ATATATAGC AGGCTGTATA
901  901  GAAGGCGGAA GAACCGGATG CGCCTGCTAT TCGCATCAAG GGACGCGATT
951  951  gaAAGAGCTG ACGGagTTGA TGTGcgaAgG aCTATGtAA AaaCGGCTTG
25  1001 CCGTTTAAAC CaTACAAGA AGAAAGCCAA GGGCAGGAAG TTCAGCAGAAG
1051 1051 CGCGCagCAA CATTCCGACA GGGCGcCAG TTGCCACATT GGGCGGAAAA
1101 1101 CGGTAGCAGA ACCTAATGTA CGATAATTG GAAGAACGCG GGAAACCGTT
1151 1151 TGAAGGAATC GGgCGGGGGC GTGGTCGGAT CGGCACAAGT A

```

This corresponds to the amino acid sequence <SEQ ID 320; ORF84>:

```

1   MAEICLIIGT PGSGKTLKMW SMMANDEMFK PDEKAIRRV FNIKGLKIP
30  51   HTYIETDAKK LPKSTDEQLS AHDMYEWIKK PENIGSIVIV DEADQVWPAR
101  101  SAGSKIPENV QWLNTHRHQG IDIFVLTOGP KLLDONLRLT VRKHYHIANS
151  151  KMGMRLLLEW KICADDPVKM ASSAFSSIIY LDKKVVLYLX XAEVHTVKNV
201  201  KRKSNVFTLP VIVLLIPFV GLSYKMLSSY GKXQEPAAQV ESAATEQQAV
35  251  LPDKTEGEPV NNGNLTADMF VFTLSEKPKS KPIYNGVRQV RTFEYIAGCI
301  301  EGGRTGCACY SHQGTALKEV TELMCKDIFVK NLFEPNPKYK ESQSQEVQQS
351  351  AQQHSDRQAV ATLGKKPKQN LMYDNWBERG KFEFGJGGGV VGSAN*

```

Further work revealed the complete nucleotide sequence <SEQ ID 321>:

```

1   ATGGCAGAGA TCTGTTTGAT AACCGGCACG CCCGGTTCAG GGAAAAACATT
40  51   AAAAAATGGTT TCCATGATGG CGAATGATGA AATGTTTAAAG CCTGATGAAA
101  101  AAGCCATACG CCGTAAAGTA TTTACGAACA TAAAGGCGCTT GAAATACCGG
151  151  CACACCTACA TAGAAACGGA CGCAAAAAGG CTGCCGAATG CGACAGATGA
201  201  CGAGCTTTCG GCGCATGATA TGTACGAATG GATAAAGAAG CCCGAAAATA
251  251  TCGGGTCTAT TGTCAATTGA GATGAAGCTC AAGACGTATG GCGGCGACGC
45  301  TCGGCAGGTT CAAAAATCCC TGAAAATGTC CAATGGCTGA ATACGCACAG
351  351  ACATCAGGGC ATTGATATAT TTGTTTGGAC TCAAGGTCCT AAGCTTCTAG
401  401  ATCAAAATCT TAGAAGCGTT GTACGGAAC ATTACACAT CGCTTCAAC
451  451  AAGATGGGTA TCGCTACGCT TTTAGATGG AARATATGCG CGGACGATCC
50  501  CGTAAAAATG GCATCAAGCG CATCTCCAG TATCTATACA CTGGATAAAA
551  551  AAGTTTATGA CTTGTACGAA TCAGCGGAAG TTCATACCGT AARATAGGTC
601  601  AAGCGGTCAA AGTGGTTTTA CACTCTGCCA GTAATAGTAT TGCTGATTCC
651  651  CGTGTCTGTC GGCCTGTCTC ATAAATGTT TCGCATCAAG GGACGCGATT
701  701  AGAAGAACC CGCAGCACA GAATCGCGGG CACACAACA GCAGGCAGTA
751  751  CTTCCGATA AAACAGAAGG CGAGCCGGTA AATACGGCA ACCTTACCGC
801  801  AGATATGTTT GTTCCGACAT TGTCCGAAA AGCCGAGATG AAGCGGATTT
851  851  ATAACGGTGT AAGCGAGTA AGAACCTTG ATATATAGC AGGCTGTATA
901  901  GAAGGCGGAA GAACCGGATG CGCCTGCTAT TCGCATCAAG GGACGCGATT
951  951  GAAAGAAGTG ACGGAGTTGA TGTGCAAGGA CTATGTAAA AACGGCTTGC
1001 1001 CGTTTAAACC ATACARAAGA GAAGGCCAAG GGCAGGAGT TCAGCAAGCG
1051 1051 GCGCAGCAAC ATTCCGACAG GGCGRAGTT GCCACATTGG GCGGAAAAACC
1101 1101 GTAGCAGAAC CTAATGTACG ATAAATTGGA AGAACCGGGG AAACCGTTTG
1151 1151 AAGGAATCGG CGGGGGCGTG GTCGGATCGG CAAACTGA

```

This corresponds to the amino acid sequence <SEQ ID 322; ORF84-1>:

```

1  MAEICLITGT PGSGKTLKMY SMMADEMFK PDENGIRRVK FTNIKGLKIP
51 HTYIETDAKK LPKSTDEQLS AHDMYEWIKK PENIGSIIVV DEAQDVWPAR
101 SAGSKI PENV QWLNTHRHQ IDIFVLTQGP KLLDQNLTL VRKHVHIASN
151 KMGMRITLLEW KICADDEVKM ASSAFSSIIY LDKKVYDLYE SAEVHTVANKV
201 KRSKWFFYLE VIVLLIFVFF GLSYKMLSSY GKQEEPAQAQ ESAATEQQAV
251 LPDKTEGEFV NNGNLTADM FVPTLSEKPE S KPIYNGVRQV RTFEYIAGCI
301 EGGRTGCACY SHQGTALKE TELMCKDYVK NGLPFPYKKE ESQOEQVQSS
351 AQQHSDRAQV ATLGKGF*QN LMYDNWEERG KPFEIGGGV VGSAN*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF84 shows 93.9% identity over a 395aa overlap with an ORF (ORF84a) from strain A of *N.*

meningitidis:

		10	20	30	40	50	60
15	orf84.pep	MAEICLITGT PGSGKTLKMY SMMADEMFK PDENKIRRVK FTNIKGLKIPHTYIETDAKK					
	orf84a	MAEICLITGT PGSGKTLKMY SMMADEMFK PDENGIRRVK FTNIKGLKIPHTYIETDAKK					
		10	20	30	40	50	60
20	orf84.pep	LPKSTDEQLSAHDMYEWIKK PENIGSIIVVDEAQDVWPARSAGSKI PENVQWLNTHRHQ					
	orf84a	LPKSTDEQLSAHDMYEWIKK PENIGSIIVVDEAQDVWPARSAGSKI PENVQWLNTHRHQ					
		70	80	90	100	110	120
25	orf84.pep	IDIFVLTQGP KLLDQNLTLVRKHVHIASN KMGMRITLLEWKICADDEVKM ASSAFSSIIY					
	orf84a	IDIFVLTQGS KLLDQNLTLVRKHVHIASN KMGMRITLLEWKICADDEVKM ASSAFSSIIY					
		130	140	150	160	170	180
30	orf84.pep	LDMKVYDLYXAEVHTVNRKRSKWFFYLEVIVLLIFVFFGLSYKMLSSY GKQEEPAQAQ					
	orf84a	LDMKVYDLYESA EVHTVNRKRSKWFFYLEVIVLLIFVFFGLSYKMLSSY GKQEEPAQAQ					
		190	200	210	220	230	240
35	orf84.pep	ESAATEQQAVLPDKTEGEFVNNGNLTADM FVPTLSEKPKSPIYNGVRQV RTFEYIAGCI					
	orf84a	ESAATEHQAVFQDKTEGEFVNNGNLTADM FVPTLSEKPKSPIYNGVRQV RTFEYIAGCV					
		250	260	270	280	290	300
40	orf84.pep	EGGRTGCACY SHQGTALKE TELMCKDYVKNGLPFPYKKE SGGQEVQSS AQQHSDRAQV					
	orf84a	EGGRTGCTCY SHQGTALKEITKEMCKDYARNGLPFPYKKE SGGQEVQSS QHSDRAQV					
		310	320	330	340	350	360
45	orf84.pep	ATLGKGPQNLMYDNWEERG KPFEIGGGV VGSANX					
	orf84a	ATLGKGPQNLMYDNWQERG KPFEIGGGV VGSANX					
		370	380	390			

The complete length ORF84a nucleotide sequence <SEQ ID 323> is:

```

1  ATGGCAGAGA TCTGTTTAT AACCGGCACG CCCGGTTCAG GGAACACATT
51 AAAAATGGTT TCCATGATGG CAACGATGA AATGTTTAA CCGGATGAAA
101 ACGGCATACG CCGTAAAGTA TTTACGAACA TCAAAAGGCTT GAAGATACCG
151 CACACCTACA TAGAAACGGA CGCGAAAAAG CTGCCCCAAT CGACAGATGA
201 GCAGCTTTCG GCGCATGATA TGTACGAATG GATAAAGAG CCCGAAATA

```


251 TCGGGTCTAT TGTCAATTGTA GATGAAGCTC AAGACGTATG GCCGGCAGCG
 301 TCGGCAGGTT CAAAAATCCC TGAAAATGTC CAATGGCTGA ATACGCACAG
 351 ACATCAGGGC ATTGATATAT TTGTTTGGAC TCAAGGCTCT AAGCTTCTAG
 401 ATCAAATGCT TAGAAGCGCT GTACGGAAC ATTACACAT CGCTCAACAC
 451 AAGATGGTA TCGCTACGCT TTGAAATGCG AAAATATGCG CGGACGATCC
 501 CGTAAAAATG GCATCAAGCG CATCTCCAG TATCTATACA CTGGATAAAA
 551 AAGTTTATGA CTTGTAAGAA TCACGCGAAG TCCATACCGT AAATAAGCTC
 601 AAGCGGTCAA AATGTTTTTA TACTCTGCCA GTAATAATAT TGCTGATTCG
 651 CGTTTTTGTG GGCCTGTCTT ATAAAATGTT AAGTAGTTAT GGAAAAAAGC
 701 AGGAAGAACC CGCAGCAGAA GAATCGCGCG CAACAGAGCA TCAGGCAGTA
 751 TTTCAGGATA AAACAGAGAG CGAGCGCGTA AACACAGGTA ACCTTACCGC
 801 AGATATGTTT GTTCCGACAT TGTCCGAAAA ACCCGAAGCG AAGCCGATTG
 851 ATAACGGTGT AAGCGAGGTA AGAACCTTTG AATATATAGC AGGCTGTGTA
 901 GAAGGCGGAA GAACCGGATG CACATGCTAT TCGCATCAAG GGACGGCATT
 951 GAAAGAAATT ACAAGGAGAA TGTGCAAGGA TTACGCAAGA AACGGATTGC
 1001 CGTTTAAACC ATATAAGAAA GAAAGCCAGG GCGCGGATGT CCAGCAAGAT
 1051 GAGCAGCACC ATTCCGACAG ACCGCAAGTT GCCACGTTGG GCGGAAAGCC
 1101 GTGGCAAAAT CTTATGTATG ATAATTGGCA GGAGCGCGGA AAACCGTTTG
 1151 AAGGAATCGG CGGGGGCGTG GTCCGATGCG CAAACTGA

This encodes a protein having amino acid sequence <SEQ ID 324>:

1 MAEICLTGT PGSGKTLKMV SMANDEMFK PDENGIRRV FTNIKGLKIP
 51 HTYIETDAKK LPKSTDEQLS AHDMYEWIKK PENIGSIVIV DEADVWFAR
 101 SAGSKIPENV QWLNTHRHQG IDIFVLTTQGS KLLDQNLRLT VRKHYHASN
 151 KMGMRITLLEW KICADDPVKM ASSAFSSIIY LDKKVVDLYE SAEVHTVNVK
 201 KRSKWFTLP VILLIPVFV GLSKMLSSY GKQEPQAAQ ESAATEHQAV
 251 FQDKTEGEPV NNGNLTADMV VPTLSEKPS KPIYNGVRQV RTFEYIAGCV
 301 EGGRTGCTCY SHQGTALKEI TKEMCKDYAR NGLFPNPKY ESQGRDVPQGS
 351 EQHHSRDPQV ATLGGKFWQN LMYDNWQERG KPFEGIGGV VGSAN*

ORF84a and ORF84-1 show 95.2% identity in 395 aa overlap:

		10	20	30	40	50	60
orf84a.pep	MAEICLTGT	PGSGKTLKMV	SMANDEMFK	PDENGIRRV	FTNIKGLKIP	HTYIETDAKK	
orf84-1	MAEICLTGT	PGSGKTLKMV	SMANDEMFK	PDENGIRRV	FTNIKGLKIP	HTYIETDAKK	
		10	20	30	40	50	60
orf84a.pep		70	80	90	100	110	120
orf84-1		70	80	90	100	110	120
		130	140	150	160	170	180
orf84a.pep	IDIFVLTTQGS	KLLDQNLRLT	VRKHYHASN	KMGMRITLLEW	KICADDPVKM	ASSAFSSIIY	
orf84-1	IDIFVLTTQGS	KLLDQNLRLT	VRKHYHASN	KMGMRITLLEW	KICADDPVKM	ASSAFSSIIY	
		130	140	150	160	170	180
orf84a.pep		190	200	210	220	230	240
orf84-1		190	200	210	220	230	240
		250	260	270	280	290	300
orf84a.pep	ESAATEHQAV	FQDKTEGEPV	NNGNLTADMV	VPTLSEKPS	KPIYNGVRQV	RTFEYIAGCV	
orf84-1	ESAATEHQAV	FQDKTEGEPV	NNGNLTADMV	VPTLSEKPS	KPIYNGVRQV	RTFEYIAGCV	
		250	260	270	280	290	300
orf84a.pep		310	320	330	340	350	360
orf84-1		310	320	330	340	350	360
		370	380	390			

```

orf84a.pep  ATLGKGKPNQNLMYDNDWQERKPFEGIGGGVVGVSANX
            |||||:|||||:|||||:|||||:|||||:|||||:
orf84-1     ATLGKGKPNQNLMYDNDWQERKPFEGIGGGVVGVSANX
            370      380      390

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF84 shows 94.2% identity over a 395aa overlap with a predicted ORF (ORF84.ng) from *N.*

gonorrhoeae:

```

10 orf84.pep  MAEICLTITGTPGSGKTLKMWSMMADEMFKPDEKAIIRKVFVTNIKGLKIPHTYIETDAKK 60
    orf84ng   MAEICLTITGTPGSGKTLKMWSMMADEMFKPDENGVRKRVFTNIKGLKIPHTHETDAKK 60

    orf84.pep  LPKSTDEQLSAHDMYEWIKKPNIGSIVIVDEAQDVPWPARSAGSKI PENVQWLNTHRHOG 120
    orf84ng   LPKSTDEQLSAHDMYEWIKKPNVGAIVIVDEAQDVPWPARSAGSKI PENVQWLNTHRHOG 120

15 orf84.pep  IDIFVLTPGPKLLDQNLRTLVRKHYHIAANKMGMRTLLEWKICADDPVKMASSAFSSIYT 180
    orf84ng   IDIFVLTPGPKLLDQNLRTLVRKHYHIAANKMGLRTLLEWKVCADDPVKMASSAFSSIYT 180

20 orf84.pep  LDKKVDYLYXXAEVHTVNKVKRSKWFFYTLPIVILLIPFVGLSYKMLSSYGGKKQEEPAAQ 240
    orf84ng   LDKKVDYLYSAEIHVTNKNVKRSKWFFYALPVIILLIPFVGLSYKMLGSYGGKKQEEPAAQ 240

25 orf84.pep  ESAATEQQAVALPDKTEGEPVNNGLNTADMFPVTLSEKPKSKPIYNGVRQVRTFEIYIAGCI 300
    orf84ng   ESAATEQQAVALPDKTEGESVNNGLNTADMFPVTLPEKPEKSPKIYNGVRQVRTFEIYIAGCI 300

30 orf84.pep  EGGRTGCACYSHQCTALKEVTELMCKDYVKNGLFNPYPKESQGGQEVQQAQHSDDRAQV 360
    orf84ng   EGGRTGCTCYSHQCTALKEVTELMCKDYVKNGLFNPYPKESQGGQEVQQAQHSDDRAQV 360

    orf84.pep  ATLGKGKPNQNLMYDNDWQERKPFEGIGGGVVGVSAN 395
    orf84ng   ATLGKGKPNQNLMYDNDWQERKPFEGIGGGVVGVSAN 395

```

The complete length ORF84ng nucleotide sequence <SEQ ID 325> is:

```

1  ATGGCAGAAA TCTGTTTGAT AACCGGCAGC CCCGGTTCAG GGAAACATT
51  AAAAATGGTT TCCATGATGG CAAACGATGA AATGTTTAAG CCAGATGAAA
101 ACGGCGTAGC CGGTAAAGTA TTTACGAACA TCAAAGGTTT GAAGATACCG
151 CACACCCACA TAGAAACAGA CGCAAGAAAG CTGCCGAATC CAACCGATGA
201 ACAGCTTTTC GCGCATGATA TGATGAATG GATCAGAAGC CTTGAAAACG
251 tcggcgCAAT CGTTATTGTC GATGAGGCC AAGACGATG GCTCCGACGC
301 TcgcCAGGTT CCAAAATCCC CGAAACGCTC CATAGCTGTA ACACACACAG
351 GCATCAGGGC ATAGATATAT TGTATTGAC ACNAGTCTCT AAATCTTAG
401 ATCAGAACCT GCGAACATTG GTTAAAAGAC ATTACACATC TCGGGCCAAC
451 AAAATGGGTT TCGCTACCCCT GCTTGAATGG AAAGTATGCG CGGATGACCC
501 GGTAAAAATG GCATCAAGTG CATTTTCCAG TATCTACACA CTGGATAAAA
551 AAGTTTATGA CTTGTACGAA TCCGACAGAA TTCACCGGT AAACAAGTGC
601 AAGCGTTCAA AATGGTTTGA TGCATTGCCC GTCATCATAT TATTGATTC
651 GCTATTTTGC GTTGTGCTTT ACAAAATGTT GGGCAGTTAC GGAAAAAAC
701 AGGAAGAACC CGCAGACAAA GAATCGGCGG CAACAGACAA GCAGGCAGTA
751 CTTCGGGATA AAACAGAAAG AGAATCGGTG AATAACGGAA ACCTTACGCG
801 AGATATGTTT GTTCCGACAT TGCCCGAAAA ACCCGAAAGC AAGCCGATT
851 ATRACGGTGT AAGGCAGGTA AGGACCTTTG AATATATAGC AGGCTGTATA
901 GAAGGCGGAA GAACCGGATG CACCTGCTAT TCGCATCAAG GCAGCGCATT
951 GAAGAAGTGT ACGGATTGTA TGTGCAAGGA CTATGTAAAA AACGCTTCG
1001 CGTTAACCC ATACAAAGAA GAAGGCCAAG GCAGCAAGT TCAGCAAGC
1051 GCGCAGCAAC ATTCCGACAG GCGCAAGTTT GCCACCTTGG CGGGAAGAAC
1101 CGACGACAA CTAATGTACG ACAAATGGGA AGAACGCGGG AAACCGTTTG
1151 AAGGAATCGG CGGGGCGGTG GTCCGATCGG CAAACTGA

```

This encodes a protein having amino acid sequence <SEQ ID 326>:

```

1  MAEICLTITGTPGSGKTLKMW SMMADEMFK PDENGVRKRV FTNIKGLKIP

```

5 51 HTHIETDAKK LFKSTDEQLS AHDMEYEWIKK PENVGAIVIV DEAQDVWPAR
 101 SAGSKIPENV QWLNTHRHGG IDIFVLTQGP KLLDQNLRTL VKRHYHIAAN
 151 KMGRLTLLEW KVCADDPVKM ASSAFSSIYT LDKKVYDLYE SAEIHTVKNKV
 201 KRSKWYALP VTILLIPLFV GLSYKMLGSY GRKQEEPPAQ ESAATQQAQV
 251 LDKKVYDLYE SAEIHTVKNKV GRKQEEPPAQ ESAATQQAQV
 301 EGGRTGCTCY SHQGTALKEV TELMCKDYVK NGLEFNPYKE ESQGEVQQAQ
 351 AQQRTSDRAQV ATLGKPKQON LMYDNWEERG KPFEGIGGGV VGSAN*

ORF84ng and ORF84-1 show 95.4% identity in 395 aa overlap:

10 orf84-1.pep 10 20 30 40 50 60
 orf84ng MAEICLITGTPGSGKTLKVMSSMANDEMFKPDENGRRKVFNTNIKGLKIPHTYIETDAKK
 15 orf84-1.pep 70 80 90 100 110 120
 orf84ng LPKSTDEQLSAHDMYEWIKK PENIGSIVIVDEAQDVWPARSAGSKI PENVQWLNTHRHGG
 20 orf84-1.pep 130 140 150 160 170 180
 orf84ng IDIFVLTQGP KLLDQNLRTL VKRHYHIAAN KMGRLTLLEWKICADDPVKM ASSAFSSIYT
 25 orf84-1.pep 190 200 210 220 230 240
 orf84ng LDKKVYDLYE SAEIHTVKNKV GRKQEEPPAQ
 30 orf84-1.pep 250 260 270 280 290 300
 orf84ng ESAATEQQAVALPDKTEGEVNMGNLTADMFVPTLSEKPEKPIYNGVRQVTFEYIAGCI
 35 orf84-1.pep 310 320 330 340 350 360
 orf84ng EGGRTGCTCY SHQGTALKEV TELMCKDYVK NGLEFNPYKE ESQGEVQQAQ
 40 orf84-1.pep 370 380 390
 orf84ng ATLGKPKQON LMYDNWEERG KPFEGIGGGV VGSANX
 45 orf84-1.pep 370 380 390
 orf84ng ATLGKPKQON LMYDNWEERG KPFEGIGGGV VGSANX

50 Based on this analysis, including the presence of a putative transmembrane domain (single-underlined) in the gonococcal protein, and a putative ATP/GTP-binding site motif A (P-loop, double-underlined), it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 39

55 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 327>:

1 GTGGTTTTCG TGAATGCCGA CAACGGGATA TTGGTTCAGG ACTTGCTCTT
 51 TGAACSTCAA CTGAAAAAAT TCCATATCGA TTTTACAAT ACGGGTATCG
 101 CGCGTGATTT CGCCAGCGAT ATTGAAGTGA CGGACAAGGC AACCGGTGAG

5

10

15

20

```

151  AAACCTCGAGC  GCACCATCCG  CGTGAACCAT  CCITTGACCT  TGCAGGGCAT
201  CACGATTPTAT  CAGGCGAGTT  TTGCCGACGG  CGGTTCGGAT  TTGCAATTCa
251  AGCGCTGGAA  TTTGGTGTAT  GCTTCGCGCG  AGCCTGTGCT  TTGAAGGCA
301  ACATCCATAC  ACCAGTTTCC  GTTGAAATTT  GGCAACACA  AATATCGTCT
351  TGAGTTCGAT  CAGTTCACCT  CTATGAATGT  GGAGACATG  AGCAGGGGG
401  CGGAACGGGA  AAAAGCCCTG  AAATCCACGC  TGCCCGATGT  CCGCGCCGTT
451  ACTCAGGAAG  GTCACAAATA  CACCAAT...  .....TACC
501  TATCCGTGAT  GCGCGAGGCC  AGGCGGTGCA  ATATAAAAAC  TATATGCTGC
551  CGGTTTTGCA  GGAACAGGAT  TATTTTTGGA  TTACCGGCAC  GCGCAGCGC.
601  TTGCAGCAGC  AATACCGCTG  GCTGCGTATC  CCCTTGGACA  AGCAGTTGAA
651  AGCGGACACC  TTTATGCGAT  TCGCTGAGTT  TTTGAAAGAT  GGGGAAGGCG
701  GCAACGCTCT  .GTTGCCGAC  GCAACCAAG  GCGCACCTGC  CGAATCCGCG
751  GAACAATTCA  TGCTGGCTGC  GGAACACACG  CTGAACATCT  TTGCACAAA
801  AGGCTATTTG  GGATTTGAGC  AATTTATTAC  GTCAATATC  CCGAAGAGC
851  AGCAGGATTA  GATGCAAGGC  TATTCTACG  AATGCTTTA  CCGCTGATG
901  AACGCTGCTT  TGGATGAAC  CAT...CCCG  TAGGCTTGG  CCGAATGCA
951  GCAGGATGAA  CGCCGGAATC  GTTTCCTGCT  GCACAGTATG  GATGCGTACA
1001  CGGTTTTCAC  CGAATATCCC  GCGCCTATGC  TGTCGAACAT  TGATGGGTTT
1051  TCGAGGTGCG  GTTCGTCGGG  TTTGCAGATG  ACCCGTTCCT  C.GGTCCGCT
1101  TTTGGTCTAT  CTC...

```

This corresponds to the amino acid sequence <SEQ ID 328; ORF88>:

25

```

1  MVFLNADNGI  LVQDLFEVFK  LKKFHIDFYN  TGMFRDFASD  IEVIDKATGE
51  KLERTIRVNH  PLTLHGHTIY  QASFADGGSD  LTFKAWNLDG  ASREPVLVLA
101  TSIHQFPLEI  GKHKYRLEFD  QFTSMNVDEM  SEGAEKREKSL  KSTLPDVRAR
151  TOEHKHYTNX  XXXXXYRIRD  APGQAVEYKN  YMLPVLQEQD  YFWITGTRSX
201  LQQQYRWLRI  PLDKQLKADT  FMALEFLKD  GEGKRRKVAD  ATKGAPEAIR
251  EQFMLAAENT  LNIFAQKGYL  GLDEFITSNi  PKEQDQKMQG  YFYEMLYGVN
301  NAALDETYXR  YGLEWQODE  ARNRFLHSM  DAYTGLTEYP  AFMLQLQDGF
351  SEVRSSGLQM  TRSXGPELLVY  L...

```

30

Further work revealed the complete nucleotide sequence <SEQ ID 329>:

35

```

1  ATGAGTAAAT  CCGCTAGATC  TCCCACACTT  CTTTCCGCTG  CGTGGTTGCG
51  TTTTTCAGC  TCCATGCGCT  TTGCAGTCAG  TTGCTGCGAT  CTGCTGCGTA
101  TTGATCCGGT  TATCGGTAAG  TGTGTGACAG  AAACACAGCC  GCACAGCGAT
151  TATTTGCTCA  AATTCGGATC  GTTTTGGCGG  CAGATTTTTG  GTTTTCTGGG
201  ACTGTATGAC  GTCATGCTTT  CGGCATGGTT  TGTGTTTATC  ATGATGTTTT
251  TGGTGGTTTC  TACCAGTTTG  TGCCGTGATC  GCAATGTGCC  GCCGTTCTGG
301  CGGAAATGAA  AGTCTTTTCG  GGAAAGAGTT  AAAGAAAAAT  CTCTGGCGGC
351  GATGCGCCAT  TCTTCGCTGT  TGAATGTAAA  AATTGCGCCC  GAGGTTGCCA
401  AACGTTATCT  GGAAGTACAA  GTTTTCAGG  GGAAAAACAT  TAACCGTGAA
451  GACGGGTCGG  TTTCTGATTG  CGCCAAAAAA  GGCAACAATG  ACAAAATGGG
501  CTATATCTTT  GCCCATGTGG  CTTTGATTGT  CATTTGCTCG  GGCGGGTTGA
551  TAGACAGTAA  CTGCTGTGTT  AAACCTGGGT  TGCTGACCCG  TCGGATTTGT
601  CGGCAACATC  AGGCGGTTTA  TGCCAAGGAT  TTCAAGCCCG  AAGATATTTT
651  GGGTGCCTCC  AATCTCTCAT  TTAGGGGCCA  CGTCAATATT  TCCGAGGGGG
701  AGAGTGCGBA  TGTGGTTTTT  CTGAAATGCC  ACAACGGGAT  ATTGGTTTCA
751  GACTTGCCCT  TTAAGTCAA  ACTGAAAAAA  TTCCATATCG  ATTTTACAAA
801  TAGGGGTATG  CCGCGTGATT  TCGCCAGGCA  TATGGAATG  ACGCACAGG
851  CACCCGGA  GAATACGAG  CGCAATCC  GCGTGAACA  TCTTTGACC
901  TTGCACGCA  TCACGATTTA  TCAGGCGAGT  TTTGCGCAGC  GCGGTTCCGA
951  TTTGACATTC  AAGCGCTGGA  ATTTGGGTGA  TGCITTCGCG  GAGCCTTGCG
1001  TGTGAAGGC  AACATCCATA  CACCAGTTTC  CGTTGGAAT  TGCCAAACAC
1051  AAATATCGTC  TTGAGTTTGA  TCAGTTCACT  TCTATGAATG  TGGAGACAT
1101  GAGCGAGGGC  GCGGAACGGG  AAAAAAGCCT  GAAATCCACG  CTGAACGATG
1151  TCCGCGCCGT  TACTCAGGAA  GGTAAAAAAT  ACACCAATAT  CGGCCCTTCC
1201  ATTTGTTTACC  GTATCCGTGA  TCGGCGAGGG  CAGGCGGTCT  AATATAAAAA
1251  CTATATGCTG  CCGGTTTTTC  AGGAACAGGA  TTATTTTTGG  ATTACGGGCA
1301  CGCGCAGCGG  CTTGCAGCAG  CAATACCGCT  GGCTGCGTAT  CCCCTTGGAC
1351  AAGCAGTTGA  AAGCGGACAC  CTTTATGGCA  TTGCGTGAGT  TTTTGAAGA
1401  TGGGGAAGGG  CGCAACGCTC  TGGTTGCCGA  CGCAACCAAA  GGGCGACCTG
1451  CGGAATCCG  CGAACATTTT  ATGCTGGCTG  GCGAAACAC  GCTGAACATC
1501  TTGCAACAAA  AAGGCTATTT  GGGATGAGC  GAATTTATTA  CTCTCAATAT
1551  CCGCGAGAG  CCGCGAGAG  AGATGAGCA  CTAATTTTAC  TGAATGCTT
1601  AGCGCTGAT  GACGCTGCTA  TTGATGATCA  CCAATAGCGG  GTAGCGCTTG
1651  CCGCAATGGC  AGCAGGATGA  AGCGCGGAAT  CGTTTCCCTG  TGCACAGTAT
1701  GGGTGCGTAC  ACGGGTTTGA  CCGAATATCC  CGCGCCTATG  CTGCTGCAAC
1751  TTGATGGGTT  TTCCGAGGTG  CGTTCGTGCG  GTTTGCAGAT  GACCCGTTCC

```

65

1801	CCGGGTGCGC	TTTTGGTCTA	TCTCGGCTCG	GTGCTGTTGG	TATTGGGTAC
1851	GGTATTGATG	TTTTATGTGC	GCGAAAAACG	GGCGTGGGTA	TTGTTTTCAG
1901	ACGCGAAAAT	CGGTTTTGCC	ATGCTTCGGG	CCCGCAGCGA	ACGGGATTTG
1951	CAGAAGGAAT	TTCAAAACA	CGTCGAGATC	CTGCAACGCG	TCGGCAAGGA
2001	CATTGATCAT	GACTAA			

This corresponds to the amino acid sequence <SEQ ID 330; ORF88-1>:

1	MSKSRRSPL	LSRPWFAFS	SMRFVALLS	LLGISASVGT	VLIQNPQSDT
51	YLVKFSFWA	QIFGFLGLD	SMRFAFVU	MLMVUVTSL	CLIRNVPPWF
101	REMKSFREX	KEKSLAMRH	SLDOKTAP	EVAKRYLVRE	GFQKTKINT
151	DGSLVLAAR	GTMMKGYAL	HYVALIVCL	GLDLSNLL	KUMLMTRIV
201	PQNVQAYAR	EKPEISLGAS	ISGNSGVNI	SEQSDASDRT	LNADNGLVQ
251	DLPEFVKLAK	FRIDELPST	PRDFASDIE	TKDQKGEK	TRIRWNPRT
301	LHGITYQPT	SNVDMDETS	KAWNLDASR	FPPVUKATSI	HGFPLEFHIG
351	KRYLEQVET	SNVDMDETS	AEREKSLST	LDNRVATOE	GKKYTNIGPS
401	YVLRDQDQ	QAVEFYKNML	PVLQEGDYF	TVTGRSGLO	YWAWRILPI
451	KOLKADTFW	FRLEKQDGS	RKLVDATAT	GAPAEITRF	MLAEANTLND
501	FAQKGYLGL	EFTSNTPEK	QDKMQGYVF	EMLYGVMAA	LEDITRYGL
551	PEWQOQEAR	RFLHSDMAY	TGTYEPAPM	LVLQDGFSE	RDSGTQMTS
601	PGALLYVGS	VLLVGLTVM	FVYREKRAW	LFSDGKIRFA	MSRSASRDL
651	QCKPEKHVS	LORLQKLDN	D*		

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF88 shows 95.7% identity over a 371aa overlap with an ORF (ORF88a) from strain A of *N. meningitidis*:

		10	20	30
orf88.pep		MVFLNADNGILVQDLPEFVKLKKFHIDFYNN 		
orf88a	AKDKFPESILGASNLFRGNVNISEQSADVFLNADNGILVQDLPFEVKLKKFHIDFYNN 210 220 230 240 250 260			
	40	50	60	70
orf88.pep	TGMPRDFASDIEVTDKATGEKLEKLTIRVNHPLTLHGIIITYQASFADGGSDLTFKAWNGLGD 			
orf88a	TGMPRDFASDIEVTDKATGEKLEKLTIRVNHPLTLHGIIITYQASFADGGSDLTFKAWNGLGD 270 280 290 300 310 320			
	100	110	120	130
orf88.pep	ASREPVVLKATSIHQFFLEIGHKHRYLEFDQFTSMNVEDMSGEAREKSLSKSTLPDVRAVV 			
orf88a	ASREPVVLKATSIHQFFLEIGHKHRYLEFDQFTSMNVEDMSGEAREKSLSKSTLNDRVAVV 330 340 350 360 370 380			
	160	170	180	190
orf88.pep	TQECHKYTNXXXXXXRIIRDAPGOVEYYNNMYLVLEVQQDYFWITGRSRXLLQQQYRWLRIRI 			
orf88a	TQECHKYTNIGPSIVYRIIRDAPGOVEYYNNMYLVLEVQQDYFWITGRSRXLLQQQYRWLRIRI 390 400 410 420 430 440			
	220	230	240	250
orf88.pep	FLDKQLKADTFMALREFLKDGEGRKRVADATKGPAEIREQFMLAAENTLNI FAQGKYLV 			
orf88a	FLDKQLKADTFMALREFLKDGEGRKRVADATKGPAEIREQFMLAAENTLNI FAQGKYLV 450 460 470 480 490 500			
	280	290	300	310
orf88.pep	GLDEFITSNIPKEQDDMKOGIFYEMLYGVMMNALDETXYRLGLEPWQQDEARNRFLHSM 			
orf88a	GLDEFITSNIPKEQDDMKOGIFYEMLYGVMMNALDETIRYRLGLEPWQQDEARNRFLHSM 510 520 530 540 550 560			
	340	350	360	370
orf88.pep	DAYTGLEYTPAMPLLQLDGFSEVRSSGLOMTRXSGLPVLVYL 			

	orf88a	DAYTGLTEYPAPMLLQLDGFSEVRS SGLQMTSRFGALLVYLGSVLLVLSTVLMFYVREKR	570	580	590	600	610	620
5	orf88a	AWVLFSDGKIRFAMSSARSERDLQKEFFKHVESLQRLKGLDNHDX	630	640	650	660	670	

The complete length ORF88a nucleotide sequence <SEQ ID 331> is:

1	ATGAGTAAAT	CCCGTAGATC	TCCCCCACTT	CTTTCGCC	CGTGGTTCGC
2	TTTTTTCAGC	TCCATCGCCT	TGCGGTCGC	TTTGCTCAGT	CTGCTGGTGA
3	TTGCATCGGT	TATCGGTACG	GTGTTCGAGC	AAAACCGACC	GCAGACGGAT
4	TATTTGGTCA	AATTCGGATC	GTTTGGGCGC	CAGATTTTGT	GTTTCTCGGG
5	ACTGTATGAC	GTCTATGCTT	CGGCATGTTT	TGTCGTTATC	ATGATGTTTT
6	TGGTGGTTTC	TACCAAGTTG	TGCCTGATTC	GCAATGTGCC	CCGCTCTCGG
7	CGCGAATATGA	AGTCTTTTTC	GGAAAAGGTT	AAAGAAAAAT	CTCTGCGCGC
8	GATGCGCCAT	TCTTCGCTGT	TGGATGTAAA	AATTCGCGCC	GAGGTGGCCA
9	AACGTTATCT	GGAAATACAA	GGTTTTCAGG	GAAAAACCAT	TACCGCTGAA
10	TACGAGTAA	CCTGCTGTTG	CGCTAAAAAA	GGCAACAATG	ACAAATGGGG
11	GACGGGTCCG	TCTTGATTTG	CTTTGATGTT	CAATTGCGTG	GCGCGGTGTA
12	CTATATCTTT	GCCCATGTTC	CTTTGATGTT	TGCTGACCGC	TCGGATGTTT
13	TACGAGTAA	CCTGCTGTTG	AACTGGGTA	TTCAAGCCCG	AAAGTATTTT
14	CGGTCGCTCC	AATCTCTCAT	TTAGGGGCGA	CGTCAATATT	TCGAGGGGCG
15	AGAGTGCAGA	TGTGGTTTTC	CTGAATGCGG	ACAAACGGAT	ATTGGTTTCA
16	GACTTGCCTT	TTGAAGTCAA	ACTGAAAAAA	TTCCATATCG	ATTTTTACAA
17	TACGGGTATG	CCGCGCGATT	TGCGCAGTGA	TATTGAAGTA	ACGGATAAAG
18	CAACCGGTGA	GAACCTCGAG	CGCACCATCC	CGGTGAACCA	TCCTTTGACC
19	TTGCACGGCA	TACCAAGTTA	TGAGGCGAGT	TTGCGCGACG	GCGGTTCCGA
20	TTTGACATTC	AAGGCGTGGG	ATTTGGGTGA	TGCTTCGCGC	GAGCCTGTGC
21	TGTTGAAGGC	AACATCCATA	CACCAATTTT	CGTTGGAAAT	TGGCAAAACAC
22	AAATATCCTC	TTGAGTTCGA	TCAGTTTACT	TCTATGAATG	TGGAGGACAT
23	GAGCGAGGGC	CGGGAACGGG	AAAAAAGCCT	GAATCCACCG	CTGAGCGATG
24	TCGCGCGCCT	TACTCAGGAA	GGTAAAAAAT	ACACCAATAT	CGCGCTTCCC
25	ATTGTTTACC	GTATCCGTGA	TGCGGCGAGG	CAGCGGTCGC	AAATGTTTGA
26	CTATATGCTG	CCGTTTGTGC	AGGAACAGGA	TATTTTGTGC	ATTACAAAAA
27	CCGCGAGGGC	CTTGCAGCAG	CAATACCGCT	GGCTGGGTAT	CCCTGTGACG
28	AAGCAGTTGA	AAGCGAGCAT	CTTTATGGCA	TGCGTGAGT	TTTTGAAGAG
29	TGGGGAAGGG	CGCAAAACGC	TGTTGCGCA	CGCAACCAA	GGCGACCTG
30	CCGAAATCCG	CGAACAAATC	ATGCTGGCTG	CGGAAACAC	GCTGAACATC
31	TTTGACAAA	AAGGCTATTT	GGGATTGGAC	GAATTTATTA	CGTCCAATAT
32	CCGGAAGAG	CAGCAGGATA	AGATGCAGGG	CTATTTCTAC	GAATGCTTTT
33	ACGCGGTGAT	GAACGCTGCT	TTGGATGAAA	CCATGCGCCG	GTACGCGTTG
34	CCGCAATGGC	AGCAGGATGA	AGCGCGGAAT	CGTTTCTGCG	TGCACAGTAT
35	GGATGCGTAC	ACGGGTTTGA	CCGAATATCC	CGCGCTATG	CTGCTGCAAC
36	TTGATGGGTT	TCCGAGGTG	CGTTTCTGCG	GTTTGAGAT	GACCGGTTCC
37	CCGCGTGGCG	TTTTGGTCTA	TCTCGGCTCG	GTGCTGTGG	TATTGGGTAT
38	GGTATTGATG	TTTTATGTGC	CGGAAAAACG	GGCGTGGGTA	TGTTGTTTCA
39	ACGGCAAAAT	CCGTTTTGCC	ATGCTCTCGG	CCCGCAGCGA	ACGGGATTGG
40	CAGAAGGAAT	TTCCAAAACA	CGTCAGAGAT	CTGCACACGG	TCGGCAAGGA
41	CTTGAATCAT	GACTGA			

This encodes a protein having amino acid sequence <SEQ ID 332>:

1	MSKRRSPPL	LSRPWFAFFS	SMRFAVALLS	LLGIASVIGT	VLQONQPTD
2	YLVKFGSEWA	QIFGFLGLYD	VYASAEVUVI	MMFLVVSTSL	CLIRNVPPFW
3	REMKSPREKV	KEKSLAAMRH	SSLDVKIAP	EVAKRYLEVO	GFQKGLINRE
4	DGSLVIAAKK	GTMMKWGYIF	AHVALIVICL	GGILDSNLLL	KLGMILTGRIV
5	PDNQAVYAKD	FKPESILGAS	NLSFRGNVNI	SEGGADVVVF	LNADNGLVLQ
6	DLPFVEVLKK	PHIDFYNTGM	PRDFASDIEV	TDKATGEKLE	RTIRVNHPLT
7	LHGTTIYQAS	FADGSSDLTF	KAWNLGDASR	EPVVLKATSI	HOFPLEIGKH
8	KYRLEFDQFT	SMNVEDMSEG	AEREKSLKST	LNDVRATQGE	GKKYTNIGPS
9	IYVIRIDAAG	QAVEYKNYML	PVLQEQDYFW	ITGTRSLGLQ	QYRWLRIPLD
10	KQLKADTFMA	LREPLKDGE	RKRIVADATK	GAPAEIREQF	MLAAENTLNI
11	FAQKGYLGLD	EFITSNIPKE	QDKMQGYFFY	EMLYGVMAAA	LDETIRRYGL
12	PEWQDQDEARN	RFLHSMDAY	TGLTEYPAPM	LLQDGFSEV	RSSGLQMTSR
13	PGALLVYLGS	VLIVLGTGLM	FVREKRAWW	LFSDGKIRFA	MSSARSERDL
14	QKEFFKHVES	LQRLKGLDNH	D*		

ORF88a and ORF88-1 100.0% identity in 671 aa overlap:

	orf88a.pep	MSKSRSPPLLSRPFFAFFSSMRFAVALLSLLGIASVIGTVLQQNQPQTDYLVKFGSFWA	60
	orf88-1	MSKSRSPPLLSRPFFAFFSSMRFAVALLSLLGIASVIGTVLQQNQPQTDYLVKFGSFWA	60
5	orf88a.pep	QIFGFLGLYDVYASAWFVIMMFLVSVTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRH	120
	orf88-1	QIFGFLGLYDVYASAWFVIMMFLVSVTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRH	120
10	orf88a.pep	SSLLDVKIAPEVAKRYLEVQGFQKGTINREDGSVLIAAKKGTMNKWGYIFAHVALIVICL	180
	orf88-1	SSLLDVKIAPEVAKRYLEVQGFQKGTINREDGSVLIAAKKGTMNKWGYIFAHVALIVICL	180
15	orf88a.pep	GGLIDSNLLKLGLMTGRIVPDNQAVYAKDFKPESILGASNLSFRGNVNISEGQSADVVF	240
	orf88-1	GGLIDSNLLKLGLMTGRIVPDNQAVYAKDFKPESILGASNLSFRGNVNISEGQSADVVF	240
	orf88a.pep	LNADNGILVQDLPEFVKLKKFHIDFYNTGMPDRFASDIEVTDKATGEKLELTIKRVNHPLT	300
	orf88-1	LNADNGILVQDLPEFVKLKKFHIDFYNTGMPDRFASDIEVTDKATGEKLELTIKRVNHPLT	300
20	orf88a.pep	LHGITIYQASFADGGSDLTFKAWNLGDASREPVVLKATSIHQFPLEIGKHKYRLEFDQFT	360
	orf88-1	LHGITIYQASFADGGSDLTFKAWNLGDASREPVVLKATSIHQFPLEIGKHKYRLEFDQFT	360
25	orf88a.pep	SMNVEDMSEGAEREKSLKSTLNDVRAVTEQGGKYTNIGPSIVYRIRDAAGQAVEYKNYML	420
	orf88-1	SMNVEDMSEGAEREKSLKSTLNDVRAVTEQGGKYTNIGPSIVYRIRDAAGQAVEYKNYML	420
30	orf88a.pep	PVLQEQQDYFWITGTRSGLQQQYRWLRIPLDKQLKADTFMALREFLKDGEGKRRLVADATK	480
	orf88-1	PVLQEQQDYFWITGTRSGLQQQYRWLRIPLDKQLKADTFMALREFLKDGEGKRRLVADATK	480
	orf88a.pep	GAPAEIREQFMIAAENTLNIFPAQKGYLGLDEITTSNIPKEQQDKMQGYFYEMLYGVMNAA	540
	orf88-1	GAPAEIREQFMIAAENTLNIFPAQKGYLGLDEITTSNIPKEQQDKMQGYFYEMLYGVMNAA	540
35	orf88a.pep	LDETIRRYGLPEWQQDEARNRFLHSDMAYTGLTEYPAPMLLQDGFSEVRSSGLQMTRS	600
	orf88-1	LDETIRRYGLPEWQQDEARNRFLHSDMAYTGLTEYPAPMLLQDGFSEVRSSGLQMTRS	600
40	orf88a.pep	PGALLVYLGSVLLVLGTVMFYVREKRAWVLSDGKIRFAMSSARSDRLQKEFPKHVES	660
	orf88-1	PGALLVYLGSVLLVLGTVMFYVREKRAWVLSDGKIRFAMSSARSDRLQKEFPKHVES	660
45	orf88a.pep	LQRLGKDLNHD 672	
	orf88-1	LQRLGKDLNHD 672	

Homology with a predicted ORF from *N.gonorrhoeae*

50 ORF88 shows 93.8% identity over a 371aa overlap with a predicted ORF (ORF88.ng) from *N. gonorrhoeae*:

	orf88.pep	MVFLNADNGILVQDLPEFVKLKKFHIDFYNTGMPDRFASDIEVTDKATGEKLELTIKRVNH	60
	orf88ng	MVFLNADNGILVQDLPEFVKLKKFHIDFYNTGMPDRFASDIEVTDKATGEKLELTIKRVNH	60
55	orf88.pep	PITLHGITIYQASFADGGSDLTFKAWNLGDASREPVVLKATSIHQFPLEIGKHKYRLEFD	120
	orf88ng	PITLHGITIYQASFADGGSDLTFKAWNLGDASREPVVLKATSIHQFPLEIGKHKYRLEFD	120
60	orf88.pep	QFTSMNVEDMSEGAEREKSLKSTLPDVRAVTEQGGKYTNXXXXXXYRIRDAAGQAVEYKN	180
	orf88ng	QFTSMNVEDMSEGAEREKSLKSTLNDVRAVTEQGGKYTNIGPSIVYRIRDAAGQAVEYKN	180
65	orf88.pep	YMLPVLQEQQDYFWITGTRSGLQQQYRWLRIPLDKQLKADTFMALREFLKDGEGKRRLVAD	240
	orf88ng	YMLPVLQDKDYFWITGTRSGLQQQYRWLRIPLDKQLKADTFMALREFLKDGEGKRRLVAD	240

66030316-0130360

5	orf88.pep	ATKGAPEIREQFMLAAENTLNIFAQKGYLGLDEFITSNIPKEQQDKMGVFFYEMLYGVM	300
	orf88.ng	ATKDAPEIREQFMLAAENTLNIFAQKGYLGLDEFITSNIPKGGQDKMGVFFYEMLYGVM	300
10	orf88.pep	NAALDETXTRYGLPEWQODEARNRFLHLSMDAYTGLTEYPAPMLQLQDGFSEVRSSGLQM	360
	orf88.ng	NAALDETXTRYGLPEWQODEARNRFLHLSMDAYTGLTEYPAPMLQLQDGFSEVRSSGLQM	360
15	orf88.pep	TRSXGPLLVL	371
	orf88.ng	TRSPGALLVYLGSVLLVLGTVMFYVPKKRAWLFSNKKIRFAMSARSERDLQKEFPKH	420

An ORF88ng nucleotide sequence <SEQ ID 333> was predicted to encode a protein having amino acid sequence <SEQ ID 334>:

15	1	MVFZINADNGM	LVQDLPFVEVK	LKKFHIDFYN	TGMPRDFASD	IEVTDKATGE
	51	KLERTIRVNH	PLTLHGITIY	QASFADGSD	LTFKAWNLRD	ASREPVVLKA
20	101	TSIHQFPLEI	GKHKYRLEFD	QFTSMNVEDM	SEGAEREKSL	KSTLNDVRAR
	151	TQEGKKYTN	GPSIVYRI	ADGAQVEYKN	YMLPILQDKD	YFWLTGTRSG
25	201	LQQQYRWLRI	PLDKQLKADT	FMALREFIKD	GEGRKKRLVAD	ATKDAPEIR
	251	EQFMLAAENT	LNIFFAQKGYL	GLDEFITSN	PKGQQDKMGV	YFYEMLYGVM
30	301	NAALDETXTRY	GLPEWQODE	ARNRFLHLSM	DAYTGLTEYP	APMLQLQDGF
	351	SEVRSSGLQM	TRSPGALLV	YLSVLLVLT	VMFYVPKKR	ARWLFNSKKI
35	401	RFAMSSARSE	RDLQKEFPKH	VESLQRLGKD	LNHD*	

Further work revealed the complete gonococcal DNA sequence <SEQ ID 335>:

25	1	ATGAGTAAAT	CCCGTATATC	TCCACACACT	CTTTCCCGTC	CGTGGTTCGC
	51	TTTTTTCAGC	TCCATCGCGT	TTGCGGTGCG	TTTGCTCAGT	CTGCTGGGTA
30	101	TTGCATCGGT	TATCGGCACG	GTGTTCACAG	AAAACACAGC	CGACAGCGTA
	151	TATTTGGTCA	AATTCGGACC	GTTTTGGACT	CGGATTTTGG	ATTTTGTGGG
35	201	TTTGTATGAT	GTCATGCTT	CGGCATGGTT	TGCTGTTATC	ATGATGTTTC
	251	TGGTGGTTTC	TACCACTTTG	TGTTTAAATC	GTAACGTTCC	GCCGTTTTCG
40	301	CGCGAAATGA	AGTCTTTCCG	GGAAAAGGTT	AAAGAAAAAT	CTCTGGCGGC
	351	GATGCGCCAT	TCTTCGCTGT	TGGATGTAAA	AATTGCCCCC	GAAGTGTCCA
45	401	AACGTTATCT	GGAGGTGGCG	GGTTTTACGG	GAATAACCGT	CAGCGGTGAG
	451	GACGGGTGCG	TTCTGATTGC	CGCCAAAAAA	GGCAcaatga	acaaATGGGG
50	501	CTATATCTTT	GCccaaagtag	ctTTGATTGT	CATTTCGCTT	GCCGCGTTGA
	551	TAGACAGTAA	CTGCTGCTG	ARGCTGGTGA	TGCTGTCGCG	TGCGATTGTT
55	601	CGCGACAATC	AGCGCGTTTA	TGCGAAGGAT	TTCAAGCCCG	AAAGTATTTT
	651	GGGTGCGCTC	AATCTCTCAT	TTACGGGCAA	CGTCAATATT	TCCGAGGGGC
60	701	AAAGTGGCGA	TGTGGTTTTC	CTGAATCCCG	ACAACCGGAT	GTGGTTTCAG
	751	GACTTGCCTT	TGAAGTCAA	ACTGAAAAAA	TTCCATATCG	ATTTTTCACAA
65	801	TACGGGTATG	CGCGCGGATT	TTGCCAGCGA	TATTGAAGTA	ACGGACAAGG
	851	CAACCGGTGA	GAACACTCGAG	CGCACCATCC	CGGTGAACCA	TCCTTTGACC
70	901	TTGCACGGCA	TCACGATTTA	TCAGCGAGAT	TTTGCCGACG	GCGGTTCCGA
	951	TTTGACATTG	AAGGCGTGGA	ATTTGAGGGA	TGCTTTCGCG	GAACCTGTGC
75	1001	TGTTGAAGGC	AACCTCCATA	CACCAGTTTC	CGTTTGGAAAT	CGGCACAAAC
	1051	AAATATCGTC	TTGAGTTCTG	TCAGTTCACT	TCATGAATG	TGGAGGACAT
80	1101	GAGCGAGGGT	CGCGAACGGG	AAAAAGCGCT	GAATCCCACT	CTGAAGCATG
	1151	TCGCGCGCGT	TACTCAGGAA	GGTAAAAAAT	ACACCAATAT	CGGCGCTTCC
85	1201	ATCGGTGACC	GCATCCGGTA	TGCGGCGSGG	CMGCGGTCG	AATATRAAAA
	1251	CTATATGCTG	CCGATTTCCT	AGGCAACAGG	CTATTTTTGG	CTGACGCGCA
90	1301	CGCGCGACGC	CTTGCAGCAG	CAATACCCCT	GCCTGGGTAT	CCCTCTGGAC
	1351	AAGCACTTGA	AAGCGGACAC	CTTTATGACA	TTGCGTGAGT	TTTTGAAGAA
95	1401	TGGGGAAGGG	CGCAACGCTG	TGGTTGCCGA	CGCAACCAAA	GACGCACCTG
	1451	CCGAAATCCG	CGAACAATTC	ATGCTGGCTG	CGGAACAACAC	GCTGAATATC
100	1501	TTTGGCGCAA	AAGGCTATTT	GGGATGGAC	GAATTTATTA	CGTCCAAATC
	1551	CCCGAAAGGG	CAGCAGGATA	AGATGCAGGG	CTATTTCTAC	GAATCTGCTT
105	1601	ACGCGGTGAT	GAACGCTGCT	TTGGATGAAA	CCATACGCCG	GTACGGCTTG
	1651	CCCGAATGGC	AGCAGGATGA	AGCCGCGAAG	CGTTTCCCTG	TGCACAGTAT
110	1701	GGATGCCATT	ACGGGGCTGA	CGGAATATCC	CGCGGCTATG	CTGCTCCAGC
	1751	TTGACGGGTT	TTTCGAGGTG	CGTTTCTCAG	GTTTTCAGAT	GACCCGTTGC
115	1801	CCGGGTGCGC	TTTTGGTGTA	Tctcggtctg	gtattgttgg	TTTTGGtgat
	1851	ggtaTttatg	tTTTATGTGC	GCGAAAAACG	GGCGTGggta	tTGTTTTTCAG
120	1901	acGGCAAAAT	CCGTTTTCCT	ATGctCTTcgg	Cgcgcagcga	ACGGGAATTTG
	1951	cAGAgggaat	TTCCAAAACA	CgtcgAGAGC	CTGCAACgpg	tcggaagga

2001 CttgaatCAT GACTga

This corresponds to the amino acid sequence <SEQ ID 336; ORF88ng-1>:

1 MSKSRISPTL LSRFWFAFFS SMRFAVALLS LLGIASVIGT VLQONQOPOTD
 5 51 YLVKFGPFWT RIFDFLGLYD VYASAWFVVI MMFLVYSTSL CLIRNVPPFW
 101 REMKSFREKV KEKSLAAMRH SSLLDVKTAP EVAKRYLEVR GFQCKTVSRE
 151 DGSVLIAAKK GTMNKGYIIF AQVALIVICL GGLIDSNLLL KLGMLAGRIQ
 201 PDNQAVYAKD FKPEISILGAS NLSFRGNVNI SEGQSAADVVF INADNGMLVQ
 251 DLPFEVKKLK FHIDFYNTGM PRDFASDIEV TDKATGEKLE RTIRVNHLPT
 301 LHGITIYQAS FADGSGDLTF KAWNLRDASR EPVVLKATSI HQFFLEIGKH
 351 KYRLEFDQFT SMNVEDMSEG AEREKSLKST LNDVRAVTOE GKKYTNIGPS
 401 IVYRIRDAAG QAVEYKYNML PILQDKDYFW LTGTRSGLQQ QYRWLRIPLD
 451 KQLKADTFMA LREFLKDGEG RKRLVADATK DAPAEIREQF MAAENTLNII
 501 FAQKGYLGLD EFTISNIFKG QQDKMQGYFY EMLYGVMNAA LDETIRRYGL
 551 PEWQQDEARN RFLHSMDAY TGLTEYPAPM LLQDGFSEV RSSGLQMTSR
 601 PGALLVYLGS VLLVLGTVMF FYVREKRAWL LFSDGKIRFA MSSARSERDL
 651 QKEFPKHVES LQRLGKDLNH D*

ORF88ng-1 and ORF88-1 show 97.0% identity in 671 aa overlap:

orf88-1.pep	MSKSRISPTLLSRFWFAFFSSMRFAVALLSLLGIASVIGTVLQONQOPOTDYLKFGSFWA	60
orf88ng-1	MSKSRISPTLLSRFWFAFFSSMRFAVALLSLLGIASVIGTVLQONQOPOTDYLKFGSFWA	60
orf88-1.pep	QIFGELGLDYVYASAWFVIMMFLVYSTSLCLIRNVPPFWREKMSFREKVKESLAAMRH	120
orf88ng-1	RIFDFLGLDYVYASAWFVIMMFLVYSTSLCLIRNVPPFWREKMSFREKVKESLAAMRH	120
orf88-1.pep	SSLLDVKIAPEVAKRYLEVQGFQGGKGTINREDGSLVIAAKKGTMNKGYIFAHVALIVICL	180
orf88ng-1	SSLLDVKIAPEVAKRYLEVQGFQGGKGTINREDGSLVIAAKKGTMNKGYIFAHVALIVICL	180
orf88-1.pep	GGLIDSNLLKLGMLTGRIVPDNQAVYAKDFKPEISILGASNLSFRGNVNISEGQSAADVVF	240
orf88ng-1	GGLIDSNLLKLGMLAGRIVPDNQAVYAKDFKPEISILGASNLSFRGNVNISEGQSAADVVF	240
orf88-1.pep	LNADNGMLVQDLPFEVKKLKEFHIDFYNTGMPRDFASDIEVTDKATGEKLERIRVNHLPT	300
orf88ng-1	LNADNGMLVQDLPFEVKKLKEFHIDFYNTGMPRDFASDIEVTDKATGEKLERIRVNHLPT	300
orf88-1.pep	LHGITIYQASFADGSGDLTFKAWNLRDASREFVVLKATSIHQFFLEIGKHKYRLEFDQFT	360
orf88ng-1	LHGITIYQASFADGSGDLTFKAWNLRDASREFVVLKATSIHQFFLEIGKHKYRLEFDQFT	360
orf88-1.pep	SMNVEDMSEGAEREKSLKSTLNDVRAVTOEGKKYTNIGPSIVYRIRDAAGQAVEYKYNML	420
orf88ng-1	SMNVEDMSEGAEREKSLKSTLNDVRAVTOEGKKYTNIGPSIVYRIRDAAGQAVEYKYNML	420
orf88-1.pep	PVLQEQDYFWITGTRSGLQQQYRWLRIPLDKQLKADTFMALREFLKDGEGRRKRLVADATK	480
orf88ng-1	PVLQEQDYFWITGTRSGLQQQYRWLRIPLDKQLKADTFMALREFLKDGEGRRKRLVADATK	480
orf88-1.pep	DAPAEIREQFMAAENTLNIIFAQKGYLGDLDEFTISNIPKEQQDKMQGYFYEMLYGVMNAA	540
orf88ng-1	DAPAEIREQFMAAENTLNIIFAQKGYLGDLDEFTISNIPKEQQDKMQGYFYEMLYGVMNAA	540
orf88-1.pep	LDETIRRYGLPEWQQDEARNRFLHSMDAYTGLTEYPAPMLLQDGFSEVRSSGLQMTSR	600
orf88ng-1	LDETIRRYGLPEWQQDEARNRFLHSMDAYTGLTEYPAPMLLQDGFSEVRSSGLQMTSR	600
orf88-1.pep	PGALLVYLGSVLLVLGTVMFYVREKRAWLFSDGKIRFAMSSARSERDLQKEFPKHVES	660
orf88ng-1	PGALLVYLGSVLLVLGTVMFYVREKRAWLFSDGKIRFAMSSARSERDLQKEFPKHVES	660
orf88-1.pep	LQRLGKDLNH 671	
orf88ng-1	LQRLGKDLNH 671	

Furthermore, ORG88ng-1 shows homology with a hypothetical protein from *Aquifex aeolicus*:

```
gi|2984296 (AE000771) hypothetical protein [Aquifex aeolicus] Length = 537
Score = 94.4 bits (231), Expect = 2e-18
Identities = 91/334 (27%), Positives = 159/334 (47%), Gaps = 59/334 (17%)

5 Query: 16 FAFSSMRFAVALLSLGLIASVIG-TVLQONQPQTDYLVKPGFPEWTRIFDFGLGYDVVYAS 74
+ F +S++ A+ ++ +LGI S++G T ++QNG YL +FG L DV+ S
Sbjct: 80 YDFLASLKLAIIFIMLVGLISMLGSTYIKQNGSFEWYLDQFGYDVGIWIKLWLDNVFHS 139

10 Query: 75 AWFVVMFVLVSTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRHSSLLDVKIAPEVAK 134
+ +++ ++ L V+ + C I+ +P W++ S +E++ + A +H + VKI P+ K
Sbjct: 140 WYYILFIVLLAVNLIFCSIKRLPRVWKQAFS-KERILKLEHAHKLPITVTKI-PDKDK 197

15 Query: 135 --RYLEVRFQGKTVSREDGSLVIAAKKGTMNKGYIFQAVALVICGLGIDSNLLKL 192
++L +GF+ V E + + A+KG ++ G +AL+VI G LID
Sbjct: 198 VLKFLKKKGFK-VFVEEENKLYVFAEKGRFSRLGVYITHIALLVIMAGALID----- 249

Query: 193 GMLAGRIVPDNQAVYAKDFKPESILGASNLISFRGNVNISEGQSADVVFNADNGMLVQDL 252
+I+G RG++ ++EG + DV+ + A+ L
Sbjct: 250 -----AIVGV-----RGLIIVAEGDTNDVMLVGAE--QKPKYL 280

Query: 253 PFEVKLKKFHIDFY---NTGMPDRFA-----SDIEVTDKATGEKLER--TIRVNHPLT 300
PF V L F I Y N + + FA SDIE+ + G K+E T++VN P
Sbjct: 281 PFVHLLIDFRIKTYAEENPNVDKRFQAQVSSYESDIEIN---GGKVEAKGTQKVNNEFD 337

25 Query: 301 LHGITIYQASFA--DGGSDLTFFKAWNLRDASRP 332
++QA++ DG S ++ + A +E
Sbjct: 338 FGRYRLFQATYGLIDGTSGMGVIVDRKKAHEDP 371
```

Based on this analysis, including the putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 40

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 337>:

```
1 ATGATGAGTA ATAAATATGGM AAAAAAAGGG TTTACATTGA TTGmGmTGAT
51 GATAGTCGTC GCGATACTCG GCATTATCAG CGTCATTGCC ATACCTTCTT
101 ATCAAAGTTA TATTGAAAAA GGCTATCAGT CCCAGCTTTA TACGGAGATG
151 GCGGTTATCA ACAAATATTTC CAAACAGTTT ATTTTGAAAA ATCCCTCTGA
201 CGATAATCAG ACCATCGAGA ACAAACGTGA AATATTGTCT TCAGGCTATA
251 AGATGAATCC GAAAATTTGCC AAAAAATATA GTGTTTCGGT AAAGTTTGTCT
301 GATAAGGAAA AATCAAGGGC ATACAGGTTG GTCGGCGTTC CGAAGGCGGG
351 GACGGGTTAT ACITTTGTGGG TATGGATGAA CAGCGTGGGC GACGATACA
401 AATGCCGTGA TGGCGGTTCT GCCCAAGCCC ATTTGGAGAC CTTGTCTCTA
451 GATGTCGGCT GTGAAGCCTT CTCTAATCGT AAAAAATA
```

This corresponds to the amino acid sequence <SEQ ID 338; ORF89>:

```
1 MMSNXMXQKG FTLIXXMIVV AILGIISVIA IPSYXSIEIK GYQSQLYTEM
51 XGINNISQKF ILKNPLDDNQ TIENKLEIFV SGYKMNPKIA KKYSVSVKVF
101 DKEKSRAYRL VGVPKAGTGY TSVWMMNSVG DGYKCRDAAS AQAHLETLS
151 DVGCEAFSNR KK*
```

Further work revealed the complete nucleotide sequence <SEQ ID 339>:

```
1 ATGATGAGTA ATAAATATGGA AAAAAAAGGG TTTACATTGA TTGAGATGAT
51 GATAGTCGTC GCGATACTCG GCATTATCAG CGTCATTGCC ATACCTTCTT
101 ATCAAAGTTA TATTGAAAAA GGCTATCAGT CCCAGCTTTA TACGGAGATG
151 GTCGGTATCA ACAAATATTTC CAAACAGTTT ATTTTGAAAA ATCCCTCTGA
```

201 CGATTAATCAG ACCAATCGGAA ACRAAATCTGGA AATATTTGTC TCAGGGCTATA
251 AGATGATGAT CAAAATATGCC AAAAARAATATA GTGTGTTTCGGT AAAGTTTGTGTC
301 GATAGAAGAAA AATCAAGGCG ATACAGGTGTG GTCCGCGCTTC CGAAGAGCGGGG
351 GACGGGTTAT ACTTTGTTCGG TATGGATGAA CAGCGCTGGGC GAGGGATATAAT
401 AATGCTGCAGT TGCGAGCTCTT GCCCAACGCT ATTTGGAGAGC CTTGTCTCTCA
451 GATGTCGGGCT GTGAAGCTCT CTCTAATCTGT AAAAAATAAA

This corresponds to the amino acid sequence <SEQ ID 340; ORF89-1>:

10

1	MMSNKMEQKG	FTLIEMMIVV	AILGIISVIA	IPSYQSYIEK	GYQSQLYTEM
51	VGINNISKQF	ILKNPLDDNQ	TIENKLEIFV	SGYKMNPKIA	KKYSVSVKFFV
101	DKEKSRAYRL	VGVPKAGTGY	TLSVWMNSVG	DGYKCRDAAS	AQAHLETLS
151	DVGCEAFSNR	KK*			

Computer analysis of this amino acid sequence gave the following results:

Homology with Pile of *N. gonorrhoeae* (accession number Z69260).

ORF89 and Pile protein show 30% aa identity in 120a overlap:

15 orf89 8 QKGFTLIXMIVVAILGIIISVIAIPSYXSYIEKGQYQSLYTEMXGINISKQFILKNPL- 66
QKGFTLI MIV+AI+GI++ +A+P+DY Y + S+ G K S + L + L +
Pile 5 QKGFTLI MIVVIAIVGILAAVLAIVQYITARAQVSAILLAEQKSAVTEYLLHGW 64

20 orf89 67 -DDNQTIENTLEIFVSGYKMPKIAKKYSVSVKFDKESRAYRLVGVPKAGTGYTLSSW 125
DN + +G + KI KY SV + GVN K G L S+
Pile 65 PKDNTS-----AGVASSDKIKGKYQSVTVKGVVTAEMASTGVNKIOGKLSW 115

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF89 shows 83.3% identity over a 162aa overlap with an ORF (ORF89a) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
	orf89.pep	MMSNXMKQGFLLIXXMIIVAILGLIISVAFPSYXSYLEGYQSQLYTEMXGINNISKQF					
30	orf89a	MMSNKMEOQGFLLIXXXXXAIXXXSVXXXXXXSYIEKGYQSQLYTEMVGINNISKQF					
		10	20	30	40	50	60
		70	80	90	100	110	120
	orf89.pep	ILKNPLDNDQTIENKLEIFVSGYKMNPKIAEKYVSVSFVDKEKSRAYRLVGVPKAGTGY					
35	orf89a	ILKNPLDNDQTIKSLKLEIFVSGYKMNPKIAEKYVNSVHFVNEEKPRAYSLVGVPKGTCTGY					
		70	80	90	100	110	120
		130	140	150	160		
	orf89.pep	TLSVWMNSVGDGYKCRDAASAQAHLTLSSDVGCEAFSNRKKXX					
40	orf89a	TLSVWMNSVGDGYKCRDAASARAHLTLSSDVGCEAFSNRKKXX					
		130	140	150	160		

The complete length ORF89a nucleotide sequence <SEO ID 341> is:

45	1	ATGATGAGTA	ATFAAAATGGA	ACAAATAAGG	TTTCATTGTA	TTTGANGNTN
	51	TATNGNCNTC	CGCATACNCN	CCGNTTACNG	CGTCATNCN	TATNNNTNCN
	101	ATNCATAGTA	TATTTGAAAAA	GGCTATTCAGT	CCCGATCTTTA	TACGGAGATG
	151	TGCGGATGTA	ACACATATTC	CAACAGCTNT	ATTITGAAAAA	ATCTCCGTGTA
	201	CGATATGATG	CGACCTCAGA	AGACCTGATG	AGACCTGATG	AGACCTGATG
	251	AGATGAACTC	GAAAAATTCG	GAATAATATG	AGTGTTCGGT	GCATTTCGTC
50	301	AATGAGAAAA	AATCAGGCGC	TATACGCTTG	TGCGGCGCTG	CRAAGACGGC
	351	GACGCGTAT	ACTTTTTCGG	TATGATGATG	CGCGCTGGCG	CGCGATGATC
	401	AATGCGTAT	CGCGCTGAT	CGCGCTGAT	CGCGCTGAT	CGCGCTGAT
	451	GATCTCGCT	GTGAGAGTCT	CTCTATATCG	AAAAAATGTA	CTTGTCCTCA

This encodes a protein having amino acid sequence <SEQ ID 342>.

55 1 MMSNKMEQKG FTLIXXXXXX AIXXXSVIX XXXYXSYIEK GYOSOLYTEM

```

51  VGINNISKQX  ILKNPLDDNQ  TIKSKLEIFV  SGYKMNPKIA  EKYNSVSHFV
101  NEEKPRAYSL  VGVPKTGTGY  TLSVWMSVSG  DGYKCRDAAS  ARAHLETLSL
151  DVGCEAFSNR  KK*

```

ORF89a and ORF89-1 show 83.3% identity in 162 aa overlap:

```

5      10      20      30      40      50      60
orf89a.pep  MMSNMKEQKGFTLIXXXXAIXXXSVIXXXYXSYIEKGYQSOLYTEMVGINNISKQX
orf89-1     MMSNMKEQKGFTLIEMMIVVAILGII SVIAIPSYQSYIEKGYQSOLYTEMVGINNISKQF
      10      20      30      40      50      60

15      70      80      90      100     110     120
orf89a.pep  ILKNPLDDNQTIKSKLEIFVSGYKMNPKIAEKYNSVSHFVNEEKPRAYSLVGVPKTGTGY
orf89-1     ILKNPLDDNQTIENKLEIFVSGYKMNPKIAKYSVSVFVDKEKSRAYRLVGVPKAGTGY
      70      80      90      100     110     120

20      130     140     150     160
orf89a.pep  TLSVWMSVSGDGYKCRDAASARAHLETLSSDVGCFAFSNRKKX
orf89-1     TLSVWMSVSGDGYKCRDAASAQALETLSSDVGCFAFSNRKKX
      130     140     150     160

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF89 shows 84.6% identity over a 162aa overlap with a predicted ORF (ORF89.ng) from *N.*

```

25  gonorrhoeae:
orf89      MMSNMXXQKGFTLIXXMIVVAILGII SVIAIPSYXSYIEKGYQSOLYTEMXGINNISKQF  60
orf89ng    MMSNMKEQKGFTLIEMMIVVTILGII SVIAIPSYQSYIEKGYQSOLYTEMVGINNVLKQF  60

30  orf89      ILKNPLDDNQTIENKLEIFVSGYKMNPKIAKYSVSVFVDKEKSRAYRLVGVPKAGTGY  120
orf89ng    ILKNPQDDNDTLKSKLKFVSGYKMNPKIAKYSVSVRFVDAEKPRAYRLVGVPNAGTGY  120

35  orf89      TLSVWMSVSGDGYKCRDAASAQALETLSSDVGCFAFSNRKK  162
orf89ng    TLSVWMSVSGDGYKCRDATSAQAYSDTLSADSGCEAFS NRKK  162

```

The complete length ORF89ng nucleotide sequence <SEQ ID 343> is:

```

40  1  aTgATGAGCA  ATAAATGGA  ACAAAAAGGG  TTTACATTGA  TTGAGATGAT
51  51  GATAGTTGTC  ACGATACTCG  GCATCATCAG  CGTCATTGCC  ATACCTTCTT
101 101  ATCAGAGTTA  TATTGAAAAA  GGCTATCAGT  COCAGCTTTA  TACGGAGATG
151 151  GTCGGTATCA  ACAATGTTCT  CAAACAGTTT  ATTTTGAARA  ATCCCGAGTA
201 201  CGATATATGAT  ACCCTCAAGA  GCRAACTGAA  AATATTTGTC  TCAGGCTATA
251 251  AGATGATATCC  GAAAAATtGCC  AAAAAATATA  GTGTTTCGGT  aagtttGTCT
301 301  gaTGCGGAAA  AACCAAGGGC  ATACAGGTTG  GTCGGCGGTC  CGAACGCGGG
351 351  GACGGGTAT  ACTTTGTGCG  TATGGATGAA  CAGCGTGGGC  GACGGATACA
401 401  AATGCCGTGA  TGCCACTTCT  GCCAGGCTCT  ATTCCGACAC  CTTGTCGGCA
451 451  GATAGCGGCT  GTGAAGCTTT  CTCTAATCGT  AAAAAATAG

```

This encodes a protein having amino acid sequence <SEQ ID 344>:

```

50  1  MMSNMKEQKG  FTLIEMMIVV  TILGIISVIA  IPSYQSYIEK  GYQSOLYTEM
51  51  VGINNVLKQF  ILKNPQDDND  TLKSKLKIFV  SGYKMNPKIA  KKYSVSVRFV
101 101  DAEKPRAYRL  VGVFNAGTGY  TLSVWMSVSG  DGYKCRDATS  AQAYSDTLSA
151 151  DSGCEAFSNR  KK*

```

This gonococcal protein has a putative leader peptide (underlined) and N-terminal methylation site (NMePhe or type-4 pili, double-underlined). In addition, ORF89ng and ORF89-1 show 88.3%

identity in 162 aa overlap:

		10	20	30	40	50	60
orf89-1.pep	MMSNKM	EQKGT	LIEMMI	VVAIL	GIISV	IAIPSY	QSYIEKGYQ
orf89ng	MMSNKM	EQKGT	LIEMMI	VVAIL	GIISV	IAIPSY	QSYIEKGYQ
		10	20	30	40	50	60
orf89-1.pep	ILKNP	LDND	QTIEN	KLEI	FVSGY	KMFKIA	KKYSVSV
orf89ng	ILKNP	QDDND	TLKSK	LKIFV	SGYKMN	KIAKKYS	VSVRFV
		70	80	90	100	110	120
orf89-1.pep	ILKNP	LDND	QTIEN	KLEI	FVSGY	KMFKIA	KKYSVSV
orf89ng	ILKNP	QDDND	TLKSK	LKIFV	SGYKMN	KIAKKYS	VSVRFV
		70	80	90	100	110	120
orf89-1.pep	TLNV	WMNSV	GDGYK	CRDA	ASAQA	HLETL	LSDDV
orf89ng	TLNV	WMNSV	GDGYK	CRDA	ASAQA	HLETL	LSDDV
		130	140	150	160		
orf89-1.pep	TLNV	WMNSV	GDGYK	CRDA	ASAQA	HLETL	LSDDV
orf89ng	TLNV	WMNSV	GDGYK	CRDA	ASAQA	HLETL	LSDDV
		130	140	150	160		

Based on this analysis, including the gonococcal motifs and the homology with the known PilE protein, it was predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF89-1 (13.6kDa) was cloned in the pGex vector and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 11A shows the results of affinity purification of the GST-fusion protein. Purified GST-fusion protein was used to immunise mice, whose sera gave a positive result in the ELISA test., confirming that ORF89-1 is a surface-exposed protein, and that it is a useful immunogen.

Example 41

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 345>:

```

1  ATGAAAAAT  CCTCCCTCAT  CAGCGCATTG  GGCATCGGTA  TTTTGAGCAT
51  CGGCATGGCA  TTTGCCGCC  CTGCCGAGCG  GGTAAGCCAA  ATCCGTCAAA
101  ACGCCACTCA  AGTATTGAGC  ATCTTAAAAA  ACGCGGATGC  CAACACCGCT
151  CGCCAAAAAG  CGGAAGCCTA  TGCATTCC  TATTTCGATT  TCCAACGTAT
201  GACCGCATTG  GCGGTGGCA  ACCCTTGGCG  TCGTCGTCC  GACGCGGCAA
251  AACACGCTT  GGCn.AGAA  TTCAACCC...

```

This corresponds to the amino acid sequence <SEQ ID 346; ORF91>:

```

1  MKKSSLISAL  GIGILSIGMA  FAFADAVSQ  IRQNTQVLS  ILKNGDANTA
51  RQKAEAYAIP  YDFQRM TAL  AVGNFWKTXS  DXKQQLAXE  FQP...

```

Further work revealed the complete nucleotide sequence <SEQ ID 347>:

```

1  ATGAAAAAT  CCTCCCTCAT  CAGCGCATTG  GGCATCGGTA  TTTTGAGCAT
51  CGGCATGGCA  TTTGCCGCC  CTGCCGAGCG  GGTAAGCCAA  ATCCGTCAAA
101  ACGCCACTCA  AGTATTGAGC  ATCTTAAAAA  ACGCGGATGC  CAACACCGCT
151  CGCCAAAAAG  CGGAAGCCTA  TGCATTCC  TATTTCGATT  TCCAACGTAT
201  GACCGCATTG  GCGGTGGCA  ACCCTTGGCG  TCGTCGTCC  GACGCGGCAA
251  AACACGCTT  GGCn.AGAA  TTCAACCC...
351  CATCTGTCAT  AAGGCGCGCA  AAGAAATCAT  GTCCGCGCG  GAACTCGGGG
401  TACCCGGGCA  AAAACCGGTC  AACATGAGAT  TCACCACCTA  CCAAGCGGCG
451  GGTAATATCC  GTACCTACAA  CGTCGCCATC  GAAGGCGGCA  GCCTGGTTAC
501  CGTGTACCGC  AACCAATTTC  GCGAAATTAT  CAAAGCGGAA  GCGCTGGAGC
551  GACTGATTGC  CGAGTTGAAA  GCCAAAAAGC  GCGGCAATAA  A

```

This corresponds to the amino acid sequence <SEQ ID 348; ORF91-1>:

```

1 MKKSSLSISAL GIGILSIGMA FAAPADAVSQ IRQNATQVLS ILKNGDANTA
51 RQKAEAYATP YDFQRM TAL AVGNPWRTAS DAQKQALAKE FQTLIRTYG
101 GTMLKLKNNAN VNVKDNPIVN KGGKEIIVRA EVGVPQGKPV NMDFTTYQSG
151 GKRYRTYNVAI EGASLVTYR NQFGEIIRAK GVDGLIAELK AKNGGK*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF91 shows 92.4% identity over a 92aa overlap with an ORF (ORF91a) from strain A of *N. meningitidis*:

```

10      10      20      30      40      50      60
orf91.pep MKKSSLSISALGIGILSIGMAFAAPADAVSQIRQNATQVLSILKNGDANTARQKAEAYATP
orf91a    MKKSSFSISALGIGILSIGMAFAAPADAVNQIRQNATQVLSILKSGDANTARQKAEAYATP
15      10      20      30      40      50      60
orf91.pep YDFQRM TALAVGNPWXTXSDXQKQALAKEFPQ
orf91a    YDFQRM TALAVGNPWRTASDAQKQALAKEFQTLIRTYSGTMLKLKNNANVNVKDNPIVN
20      70      80      90      100     110     120
orf91a    KGGKEIIVRAEVGVPQGKPVNMDFTTYQSGGKYRTYNVAIEGASLVTYRNRQFGEIIRAK
130     140     150     160     170     180

```

The complete length ORF91a nucleotide sequence <SEQ ID 349> is:

```

25      1 ATGAAAAAAT CCTCCTTCAT CAGCGCATTG GGCATCGGTA TTTGAGCAT
51 CGGCATGGCA TTTGCCGCC CTGCCGACGC GGTAAACCAA ATCCGTCAAA
101 ACGCCTACTA AGTATTGAGC ATCTTAAAAA CGCGTGATGC CAACACCGCC
151 CGCCAAAAG CCGAAGCCTA TGGCATTCCC TATTTCGATT TCCACGTAT
201 GACCGCATTG GCGGTCCGCA ACCCTTGGCG CACCGGTGCC GACGCGCAA
301 AACACAGGCTT GGCCAAAGAA TTTCAAACCC TGCTGATCCG CACCTATTC
351 GGCACGATGC TGAATTTAAA AAACGCCAAC GTCAACGTCA AAGACAATCC
401 TACCCGGGCA AAAACCCGTC AACATGGACT TCACCACCTA CCAAGCGCG
451 GGTAAATACC GTACCTACAA CGTCGCCATC GAAGGCGCGA GCCTGGTTAC
501 CGGTACCGC AACCAATTGC GCGAAATTAT CAAGCGAAR GCGGTGGACG
551 GACTGATTGC CGAGTTGAAG GCTAAAACG GCAGCAAGTA A

```

This encodes a protein having amino acid sequence <SEQ ID 350>:

```

1 MKKSSFSISAL GIGILSIGMA FAAPADAVNQ IRQNATQVLS ILKSGDANTA
51 RQKAEAYATP YDFQRM TAL AVGNPWRTAS DAQKQALAKE FQTLIRTYG
101 GTMLKLKNNAN VNVKDNPIVN KGGKEIIVRA EVGVPQGKPV NMDFTTYQSG
151 GKRYRTYNVAI EGASLVTYR NQFGEIIRAK GVDGLIAELK AKNGSK*

```

ORF91a and ORF91-1 show 98.0% identity in 196 aa overlap:

```

45      10      20      30      40      50      60
orf91a.pep MKKSSFSISALGIGILSIGMAFAAPADAVNQIRQNATQVLSILKSGDANTARQKAEAYATP
orf91-1    MKKSSLSISALGIGILSIGMAFAAPADAVSQIRQNATQVLSILKNGDANTARQKAEAYATP
50      10      20      30      40      50      60
orf91a.pep YDFQRM TALAVGNPWRTASDAQKQALAKEFQTLIRTYSGTMLKLKNNANVNVKDNPIVN
orf91-1    YDFQRM TALAVGNPWRTASDAQKQALAKEFQTLIRTYSGTMLKLKNNANVNVKDNPIVN
55      70      80      90      100     110     120
orf91a.pep KGGKEIIVRAEVGVPQGKPVNMDFTTYQSGGKYRTYNVAIEGASLVTYRNRQFGEIIRAK
130     140     150     160     170     180

```

```

      |||
orf91-1  KGGKEIIVRAEVGVPQGKPVNMDTTYQSGGKYRTYNVAIEGASLTVYVRNQFGEI IKAK
      130      140      150      160      170      180

5      orf91a.pep  GVDGLI AELKAKNGSKX
      |||
orf91-1  GVDGLI AELKAKNGGKX
      190

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF91 shows 84.8% identity over a 92aa overlap with a predicted ORF (ORF91.ng) from *N. gonorrhoeae*:

```

orf91.pep  MKKSSLSISALGIGILSIGMAFASPADAVGQIRQNATQVLTILKNGDANTARQKAEAYAIP 60
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
orf91ng    VKKSSFSISALGIGILSIGMAFASPADAVGQIRQNATQVLTILKSGDAASARPKAEAYAVP 60

orf91.pep  YDFQRM TALAVGNPWXTSDXQKQALAKEFQP 93
||||:||||:||||:||||:||||:||||:||||:
orf91ng    YDFQRM TALAVGNPWRTASDAQKQALAKEFQTLLIRTYSGTMLKFNATVNVKDNPIVN 120

```

The complete length ORF91ng nucleotide sequence <SEQ ID 351> is predicted to encode a protein having amino acid sequence <SEQ ID 352>:

```

1  VKKSSFSISAL GIGILSIGMA FASPADAVGQ IRQNATQVLT ILKSGDAASA
51  RPKAEAYAVP YDFQRM TAL AVGNPWRTAS DAQKQALAKE FQTLLIRTYG
25  101 GTMLKFNAT VNVKDNPIVN KGGKEIVLRA EVGIPGQKPV NMDTTYQSG
151 GKYRTYNVAI EGTSLVTYVR NQFGEI IKAK GIDGLIAELK AKNGGK*

```

Further work revealed the complete nucleotide sequence <SEQ ID 353>:

```

1  ATGAAAAAAT CCTCTTCAT CAGCGCATTG GGCATCGGTA TTTTGAGCAT
30  51 CGGCATGGCA TTTGCCCTCC CGCGCGAGCG AGTGGGACAA ATCCGCCAAA
101 ACGCCACACA GGTTTTGACC ATCTCCAAAA CGCGGACGCG GGCTCTCTGA
151 CGCCCAAAAG CCGAAGCGTA TCCGCTTCCC TATTTGATT TCACAGTAT
201 GACCATTTG CGGTCCGGCA ACCCTTGGCG TACCGGCTCC GACGCGCAAA
251 AACACAGCTT GGCACAAAGAA TTTCAAACCC TGCTGATCCG CACCTATCC
301 GGACAGATGC TGAATTCAA AAACGCGACG GTCACGTC AAGACATCC
35  351 CATCGTCAAT AAGGCGCGCA AGGAATTCGT CGTCCGTGCC GAAGTCGGCA
401 TCCCGGTGCA GAAGCCCGTC AATATGGACT TTACACCTA CCAAAGCGCG
451 GGCAATATCC GTACCTACAA CGTCGCATC GAAGGCAGA GCCTGGTTAC
501 CGGTACCGC AACCAATTCG GCGAATCAT CAAAGCCAAA GGCATCGACG
551 GGCTGATTGC CGAGTTGAAA GCCAAAACG CGGCAAAATA A

```

This corresponds to the amino acid sequence <SEQ ID 354; ORF91ng-1>:

```

1  MKKSSFSISAL GIGILSIGMA FASPADAVGQ IRQNATQVLT ILKSGDAASA
51  RPKAEAYAVP YDFQRM TAL AVGNPWRTAS DAQKQALAKE FQTLLIRTYG
101  GTMLKFNAT VNVKDNPIVN KGGKEIVLRA EVGIPGQKPV NMDTTYQSG
151  GKYRTYNVAI EGTSLVTYVR NQFGEI IKAK GIDGLIAELK AKNGGK*

```

ORF91ng-1 and ORF91-1 show 92.3% identity in 196 aa overlap:

```

      10      20      30      40      50      60
orf91-1.pep  MKKSSLSISALGIGILSIGMAFASPADAVGQIRQNATQVLTILKNGDANTARQKAEAYAIP
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
orf91ng-1    MKKSSFSISALGIGILSIGMAFASPADAVGQIRQNATQVLTILKSGDAASARPKAEAYAVP
      10      20      30      40      50      60

      70      80      90      100     110     120
orf91-1.pep  YDFQRM TALAVGNPWRTASDAQKQALAKEFQTLLIRTYSGTMLKFNATVNVKDNPIVN
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
orf91ng-1    YDFQRM TALAVGNPWRTASDAQKQALAKEFQTLLIRTYSGTMLKFNATVNVKDNPIVN
      70      80      90      100     110     120

```

5

orf91-1.pep	130	140	150	160	170	180
	KGGKEIVRAE	VGPGKPVN	MDFTYQSGG	KYRTYNVA	IEGASIVTV	YRNQFGEIIKAK
orf91ng-1						
	KGGKEIVRAE	VGIPGOKPVN	MDFTYQSGG	KYRTYNVA	IEGTSLVTV	YRNQFGEIIKAK
	130	140	150	160	170	180

10

orf91-1.pep	190
	GVDGLIAELKAKNGGKX
orf91ng-1	
	GIDGLIAELKAKNGGKX
	190

In addition, ORF91ng-1 shows homology to a hypothetical *E.coli* protein:

```

15      sp|P45390|YRBC_ECOLI_HYPOTHETICAL 24.0 KD PROTEIN IN MURA-RPON INTERGENIC
      REGION PRECURSOR (F211) >g1|606130 (U18997) ORF_f211 [Escherichia coli]
      >gi|1789583 (AE000399) hypothetical 24.0 kD protein in murZ-rpoN intergenic
      region [Escherichia coli]Length = 211

20      Score = 70.6 bits (170), Expect = 6e-12
      Identities = 42/137 (30%), Positives = 76/137 (54%), Gaps = 6/137 (4%)

Query: 59  VPFYDFQRMTALVAGNPWRTASDAQKQALAKEFQTLIRITYSGTMLKFNKATVNVKDNPI 118
      +PY + AL+G ++A+ AQA+A E+ L + Y + + + T +
Sbjct: 65  LPPVQVYKAGALVLGQYKKSATPAQRQAYFAAFREYLKQAYGALAMHQCQYQIA--PE 122

25  Query: 119 VNKGGKEIV-VRAEVGIP-GQKPVNMDFTTYQSG--GKYRTYNVAIEGTSLVTVYRNQPG 174
      G K IV +R + P G+ PV +DF ++ G ++ Y++ EG S+T+ +N+G
Sbjct: 123  QPLGDKTIVIRIVTIDNPNGPPVRLDFQWRKNQSGTGNWQAYDMIAEGVSMITTKQNEG 182

30  Query: 175 EIIKAKGIDGLIAELKA 191
      +++ KGIDGL A+LK+
Sbjct: 183  TLLRRTKGIDGLTAQLKS 199

```

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 42

The following DNA sequence was identified in *N.meningitidis* <SEQ ID 355>:

	1	ATGAACACCA	TACTCTCCCT	GATTCGCCGA	TCCGCATCTG	GCATTTCAC
40	51	CGCTTCGGCA	CATCTCGCCA	CGGCAACCTG	CATCACTAAT	GAACACGGTA
	101	TGATCAACGA	TACCTCTATC	TCAAAATAGA	GTTTTCGgnn	nnnnnnnnnn
	151	nnnnnnnnnn	nnGCCGATTC	AGCAAAAGCG	TTGCGCATGT	nnnnnnnnnn
	201	CGACCTGGCA	CGCCACGAC	CGCCAAATCG	nnnnnnnnnn	nnnnnnnnnn
	251	AAGTCACTGT	CTTGGGACGC	CCCAAGGACG	CGACGCGCTG	GATGTGCAAT
45	301	GACCGCCGTG	TCGCGCTGCA	CTGCGCTCTA	CGGCTCTCTG	TACCGAACAA
	351	CGGCGCGCAA	GAGGAGCGCG	CTCTATACCA	TACGCGCGCG	CTCATCGCGT
	401	CGGCTCTGAC	CTGCTCTGAC	CGGCGGCGCG	CACTTTTCGG	AAACGCGGAA
	451	GAACCTGATC	GTAAACACGT	AGGCGCAATA		

This corresponds to the amino acid sequence <SEQ ID 356; ORF97>:

50

1	MKHILPLIAA	SALCISTASA	HPASEPSTQN	ETAMITHTLI	SKYSFGXXXX
51	XXXXAIKSKG	MDIFAVIDHQ	EAARRNGLTM	QPAKVIVFQT	PKAGTPLMVK
101	DPAFALQLPL	RVLVTETDGG	VRAAYTDTA	LIAGSRIGED	EVANTLANAE
151	KLIQKTVGE*				

Further work revealed the complete nucleotide sequence <SEQ ID 357>:

55 1 ATGAAACACA TACTCCCCCT GATTGCCGCA TCCGCACTCT GCATTTCAAC
51 CGCTTCGGCA CATCCTGCCA GCGAACCCTC CACCCAAAAC GAAACCCGCTA

5

```

101 TGACCACGCA TACCTCACC TCAAAATACA GTTTTGACGA AACCGTCAGC
151 CGCCTTGAAG CGCCCATAAA AAGCAAGGG ATGGACATT TTGCGTCAT
201 CGACCATCAG GAAGCCGCC GCGAAACGG CTTAACGATG CAGCGCGCAA
251 AAGTCATCGT CTTCCGCCAG CCCAAAGCG GCACGCGCGT GATGGTCAAA
301 GACCCCGCCT TCGCCCTGCA ACTGCCCTCA CGCGTCTCG TTACCGAAAC
351 GGACGCGCAA GTACGCGCG CTTATACCGA TACGCGCGCC CTCATCGCGC
401 CGAGCCGCAT CGGTTTCGAC AAGTGGCAA ACACTTGGC AAACGCCGAA
451 AAATGTATAC AAAAACCGT AGCGAATAA

```

This corresponds to the amino acid sequence <SEQ ID 358; ORF97-1>:

10

```

1 MKHILPLIAA SALCISTASA HPASEPSTQN ETAMTHTLT SKYSFDETVS
51 RLETAIKSKG MDIFAVIDHQ EAARRNGLTM QPAKVIVFGT PKAGTPLMVK
101 DPAPALQLPL RVLVTETDGL VRAAYTDTRA LIAGSRIGFD EVANTLANAE
151 KLIQKTIGE*

```

Computer analysis of this amino acid sequence gave the following results:

15 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF97 shows 88.7% identity over a 159aa overlap with an ORF (ORF97a) from strain A of *N.*

meningitidis:

20

```

      10      20      30      40      50      60
orf97.pep MKHILPLIAASALCISTASHPASEPSTQNETAMTHTLT SKYSFGXXXXXXAIKSKG
          ||||| ||||| ||||| ||||| ||||| : |||||
orf97a    MKHILPLXXASALCISTASXHPASEPOTQNETAMTHTLT SKYSFDETVSRLETAIKSKG
          10      20      30      40      50      60

      70      80      90      100     110     120
orf97.pep MDIFAVIDHQEAARRNGLTMQPAKVIVFGTPKAGTPLMVKDPAPALQLPLRVLTETDGL
          ||||| ||||| ||||| ||||| ||||| |||||
orf97a    MDIFAVIDHQEAARRNGLTMQPAKVIVFGTPKAGTPLMVKDPAPALQLPLRVXVLTETDGL
          70      80      90      100     110     120

      130     140     150     160
orf97.pep VRAAYTDTRALIAGSRIGFDEVANTLANAEKLIQKTIGEX
          ||||| ||||| ||||| ||||| ||||| |||||
orf97a    VRAAYTDTRALIAGSRIGFDEVANTLANAEKLIQKTIGEX
          130     140     150     160

```

35 The complete length ORF97a nucleotide sequence <SEQ ID 359> is:

40

```

1 ATGANACACA TACTCCCCCT GANTGNCGA TCCGCACTCT GCATTCTAAC
51 CGCTTCGGNN CATCTTGCCA GCGAACCGCA AACCCAAAC GAACCGCTCA
101 TGACCACGCA TACCTCACC TCAAAATACA GTTTTGACGA AACCGTCAGC
151 CGCCTTGAAG CGCCCATAAA AAGCAAGGG ATGGACATT TTGCGTCAT
201 CGACCATCAG GAAGCCGCC GCGAAACGG CTTAACGATG CAGCGCGCAA
251 AAGTCATCGT CTTCCGCCAG CCCAAAGCG GCACGCGCGT GATGGTCAAA
301 GACCCCGCCT TCGCCCTGCA ACTGCCCTCA CGCGTCTCG TTACCGAAAC
351 GGACGCGCAA GTACGCGCG CTTATACCGA TACGCGCGCC CTCATCGCGC
401 CGAGCCGCAT CGGTTTCGAC AAGTGGCAA ACACTTGGC AAACGCCGAA
451 AAATGTATAC AAAAACCAT AGCGAATAA

```

This encodes a protein having amino acid sequence <SEQ ID 360>:

50

```

1 MKHILPLXXA SALCISTASX HPASEPOTQN ETAMTHTLT SKYSFDETVS
51 RLETAIKSKG MDIFAVIDHQ EAARRNGLTM QPAKVIVFGT PKAGTPLMVK
101 DPAPALQLPL RVXVLTETDG VRAAYTDTRA LIAGSRIGFD EVANTLANAE
151 KLIQKTIGE*

```

ORF97a and ORF97-1 show 95.6% identity in 159 aa overlap:

55

```

      10      20      30      40      50      60
orf97a.pep MKHILPLXXASALCISTASXHPASEPOTQNETAMTHTLT SKYSFDETVSRLETAIKSKG
          ||||| ||||| ||||| ||||| ||||| |||||
orf97-1    MKHILPLIAASALCISTASHPASEPSTQNETAMTHTLT SKYSFDETVSRLETAIKSKG

```

		10	20	30	40	50	60
5	orf97a.pep	MDIFAVIDHQEAARRNGLTMQPAKVIVFGT	PKAGTFLMVKDPAFALQLPLRVVXTETD	GK			
	orf97-1	MDIFAVIDHQEAARRNGLTMQPAKVIVFGT	PKAGTFLMVKDPAFALQLPLRVLTETD	GK			
		70	80	90	100	110	120
10	orf97a.pep	VRAAYTDTRALIAGSRIGFDEVANTLANAEKLIQKTIGEX					
	orf97-1	VRAAYTDTRALIAGSRIGFDEVANTLANAEKLIQKTIGEX					
		130	140	150	160		
		130	140	150	160		

15 Homology with a predicted ORF from *N.gonorrhoeae*

ORF97 shows 88.1% identity over a 159aa overlap with a predicted ORF (ORF97.ng) from *N.gonorrhoeae*:

	orf97.pep	MKHILPLIAASALCISTASAHASEPSTQNETAMITHTLISKYSFGXXXXXXAISKSG	60
20	orf97ng	MKHILPLIAASALCISTASAHASEPSTQNETAMITHTLISKYSFDETVSRLETAISKSG	60
	orf97.pep	MDIFAVIDHQEAARRNGLTMQPAKVIVFGTPKAGTFLMVKDPAFALQLPLRVLVTETDGG	120
	orf97ng	MDIFAVIDHQEAARRNGLTMQPAKVIVFGTPKAGTFLMVKDPAFALQLPLRVLVTETDGG	120
25	orf97.pep	VRAAYTDTRALIAGSRIGFDEVANTLANAEKLIQKTGVE	159
	orf97ng	VRTAYTDTRALIAGSRIGFDEVANTLANAEKLIQKTGVE	159

The complete length ORF97ng nucleotide sequence <SEQ ID 361> is predicted to encode a protein having amino acid sequence <SEQ ID 362>:

30	1	MKHILPLIAA	SAFCISTASA	HPAGKPPTQN	ETAMTHTLT	SKYSFDETVS
	51	RLETAISKSG	MDIFAVIDHQ	EAARRNGLTM	QPAKVIVFGT	PKAGTFLMVK
	101	DPAFALQLPL	RVLTETD	GK	VRTAYTDTRA	LIVGSRISFD
	151	KLIQKTIGEX*				EVANTLANAE

35 Further work revealed the complete nucleotide sequence <SEQ ID 363>:

40	1	ATGAACACA	TACTCCcct	gatcgccgca	TccgcaactCT	GCATTTC AAC
	51	CGCTTCGGCA	CACCCTGCGG	GCAAAACGCC	CACCCAAAC	GAACCGCTA
	101	TGACCAACGCA	CACCCTCACC	TCGAAATACA	GTTTTCGACGA	AACCGTCAGC
	151	CGCCTTGAAA	CGGCCATAAA	AAGCAAGGG	ATGGACATT	TTGCGTCAGC
	201	CGACCATCAG	GAAGCGGCAC	GCCGAAACGG	CCTGACCATG	CAGCGCGCAA
	251	AAGTCATCGT	CTTCGGCAGC	CCCAAGCCCG	GTAGCGcct	GATGGTCAAA
	301	GACCCCGCCT	TCGCCCTGCA	ACTGCCCTCG	CGCGCTCTCG	TTACCGAAAC
	351	GGACGGCAAA	GTACGCACCG	CCTATACCGA	TACCGCGCCG	CTCATCGTGG
	401	GCAGCGGCAT	CAGTTTCGAC	GAAGTGGCAA	ACACTTTGGC	AAACGCGCGA
45	451	AAACTGATAC	AAAAAACCGT	AGGGGAATAA		

This corresponds to the amino acid sequence <SEQ ID 364; ORF97ng-1>:

50	1	MKHILPLIAA	SALCISTASA	HPAGKPPTQN	ETAMTHTLT	SKYSFDETVS
	51	RLETAISKSG	MDIFAVIDHQ	EAARRNGLTM	QPAKVIVFGT	PKAGTFLMVK
	101	DPAFALQLPL	RVLTETD	GK	VRTAYTDTRA	LIVGSRISFD
	151	KLIQKTIGEX*				EVANTLANAE

ORF97ng-1 and ORF97-1 show 96.2% identity in 159 aa overlap:

55	orf97-1.pep	MKHILPLIAASALCISTASAHASEPSTQNETAMTHTLTSKYSFDETVSRLETAISKSG					
	orf97ng-1	MKHILPLIAASALCISTASAHASEPSTQNETAMTHTLTSKYSFDETVSRLETAISKSG					
		10	20	30	40	50	60
		10	20	30	40	50	60

		70	80	90	100	110	120
orf97-1.pep	MDIFAVIDHCEAARRNGLTMQPAKVIIVFGT	PKAGTPLMVKDPAFALQLPLRVLTETD	GK				
orf97ng-1	MDIFAVIDHCEAARRNGLTMQPAKVIIVFGT	PKAGTPLMVKDPAFALQLPLRVLTETD	GK				
		70	80	90	100	110	120
orf97-1.pep	VRAAYTDTRALIAGSRIGFDEVANTLANAEKLIQKT	VGEX					
orf97ng-1	VRTAYTDTRALIVGSRI:SFDEVANTLANAEKLIQKT	VGEX					
		130	140	150	160		
		130	140	150	160		

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF97-1 (15.3kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figures 12A & 12B show, respectively, the results of affinity purification of the GST-fusion and His-fusion proteins. Purified GST-fusion protein was used to immunise mice, whose sera were used for Western Blot (Figure 12C), ELISA (positive result), and FACS analysis (Figure 12D). These experiments confirm that ORF97-1 is a surface-exposed protein, and that it is a useful immunogen.

Figure 12E shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF97-1.

Example 43

The following DNA, believed to be complete, sequence was identified in *N.meningitidis* <SEQ ID 365>:

```

1  ATGGCTTTTA  TTACGCGCTT  ATTCAAAGC  AGTAAATGCG  TGATTGTGCC
51  GCTGATGCTC  CCGGCTTTC  AGAATGTGGC  GCGGAGGGGG  ATAGATGTGA
101  GCCGTGCGCA  AGCGAGGATA  ACCGACGGCG  GGCAGCTTTC  CATCAGCAGC
151  CGCTTCCAAA  CCGAGCTGCC  CGACCAGCTC  CAACAGGCGT  TGCGCCGGGG
201  CGTGCGCGTC  AACTTTACTT  TAAGCTGGCA  GCTTTCGGCC  CCGATAATCG
251  CTCTTATCG  GTTTAAATTG  GSGCACTGA  TTGCGCATGA  CGCAATATT
301  GACTACAAAC  TGAATTCCA  TCGCTGAC  AAGCGTACC  GCGTTACCG
351  CGGCGGCTT  TCGACAGACT  ACGACACTT  GCATGCGCA  TTGCGCGCA
401  CCGGCGCGGT  TGCCAACCTG  AACTCTCTGA  ACAGGCGCC  GCTGTCCGGT
451  GCGGAAGCAG  GGAACACAA  GCGGAAATC  CCGCTGAGC  TGTCACCTTC
501  AAAACTGCCC  AAGCCTTTC  AAATCAATGC  ATTGACTTCT  CAARACTGGC
551  ATTTGGATTC  GGGTTGAAA  CCTTAAACA  TCATCGGAA  CAAATAA

```

This corresponds to the amino acid sequence <SEQ ID 366; ORF106>:

```

1  MAFITRLFKS  SKWLIVPLML  PAFQNVAAEG  IDVSRAEARI  TDGQGLSISS
51  RFQTELPDQL  QQALRRGVPL  NFTLSWQLSA  PIASVRFKL  GQLIGDDNNI
101  DYKLSFHELT  KRYRVTVGAF  STDYDTLDA  LRATGAVANW  KVLNKGALSG
151  AEAGETKAEI  RLTLSTSKLP  KFFQINALTS  QNWLDSGWK  PLNIINKN*

```

Further work revealed the following DNA sequence <SEQ ID 367>:

```

1  ATGGCTTTTA  TTACGCGCTT  ATTCAAAGC  AGTAAATGCG  TGATTGTGCC
51  GCTGATGCTC  CCGGCTTTC  AGAATGTGGC  GCGGAGGGGG  ATAGATGTGA

```

101 GCGGTGCCGA AGCGAGGATA ACCGACGGCG GGCAGCTTTC CATCAGCAGC
 151 CGCTTCCCAA AGGAGCTGGC CGACCAGCTC CAACAGCGGT TGGCGCGGGG
 201 CGGCGCGCTC AACTTTACCT TAAGCTGGCA GCTTCCGCC CCGATAATCG
 251 CTTCCTATCG GTTTAAATTG GGGCAACTGA TTGGCGATGA CGACAATATT
 301 GACTACAAAC TGAGTTTCCA TCCGCTGACC AACCGCTACC GCGTACCGT
 351 CGGCGCGTTT TCGACAGACT ACGACACCTT GGATGCGGCA TTGCGCGCGA
 401 CGGCGCGGTT TCGCAACTGG AAGTCCCTGA ACAAGGCGCG GCTGTCGGT
 451 GCGGAAGCAG GGGAAACCAA GGCAGAAATC CGCCTGACGC TGTCCAATTCT
 501 AAAAAGTGGC AAGCGTTTTC AAATCAATGC ATTGACTTCT CAAAACCTGGC
 551 ATTTGGATTG GGGTTGAAA CCTCTAAACA TCATCGGGAA CAAATAA

This corresponds to the amino acid sequence <SEQ ID 368; ORF106-1>:

1 MAFITRLFKS SKWLIVPLML PAFQNVAAEG IDVSRAEARI TDGGQLSISS
 51 RFQTELPDQL QQALRRGVPL NFTLSWQLSA PIITASRYFKL GQLIGDDNDNI
 101 DYKLSFHLPT MRYRYTVGAF STDYDTLDA LRTATGAVANW KVLNKGALSG
 151 AEAGETKAEI RLTLSTSKLP KPFQINALTS QNWLDSGWK PLNIIGNK*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF106 shows 87.4% identity over a 199aa overlap with an ORF (ORF106a) from strain A of *N. meningitidis*:

		10	20	30	40	50	59
orf106.pep	MAFITRLFKSSK-WLIVPLMLPAFQNVAAEGIDVSRAEARITDGGQLSISSRFQTELPDQ						
orf106a	MAFITRLFKSIQWLVLLPMLSVLPDAAEGIDVSRAEARIXDGGQLSISSRFQTELPDQ						
		10	20	30	40	50	60
25		60	70	80	90	100	110
orf106.pep	LQQALRRGVPLNFTLSWQLSAPIIIASRYFKLGLIGDDNDNDYKLSFHLPTKRYRYTVGA						
orf106a	LQXAXXRGVXLNXTLXWQLSAPIIIASRYFKLGLIGDDNDXIDYKLSFHLPTNRYRYTVGA						
30		70	80	90	100	110	120
orf106.pep	FSTDYDTLDAALRATGAVANWVNLNKGALSGAEAGETKAEIRLTSTLSKLPKPFQINALT						
orf106a	FSTXYDLDAALRATGAVANWVNLNKGALSGAEAGETKAEIRLTSTLSKLPKPFQINALT						
35		120	130	140	150	160	170
orf106.pep							
orf106a							
		130	140	150	160	170	180
40		180	190	199			
orf106.pep	SQNWLDSGWKPLNIIGNKX						
orf106a	SQNWLDSGWKPLNIIGNKX						
		190	200				

Due to the K→N substitution at residue 111, the homology between ORF106a and ORF106-1 is 87.9% over the same 199 aa overlap.

45 The complete length ORF106a nucleotide sequence <SEQ ID 369> is:

1 ATGGCTTTTA TTACGCGCTT ATTCAAAGC ATTAACAATG GGCTTGTCCT
 51 GCTGCCGATG CTTTCCGTTT TGCCGGAGCG GCGCGCGGAG GGGATAGATG
 101 TGAGCCGCGC CGAAGCGAGG ATAANCGACG GCGGCGAGCT TTCCATNAGN
 151 AGCCGCTGCC AAACCGACTT GCCGACACAG CTCACANNNG GGNNGGCGCG
 201 GCGGCTGNCG CTCACCTNTA CTTAAGATNG CCGCTTTCG GCGCGATATA
 251 TCAGCTCTTA TCGGCTTNA TGGGCGCAAC TCATTGGGCA GCACACNAT
 301 ATTGACTACA AACTGAGTIT CCATCCGCTG ACCAACCGCT ACCGCTGATG
 351 CGTCCGCGCG CTTTCGACAG ANTACGACAC CTTGGATGCG GCATTGCGCG
 401 CGACCGGCGC GGTTCGCAAC TGGAAAGTCC TGAACAAAG CCGCTGTCC
 451 GGTGCGGAAG CAGGGGAAC CAAGCGGAA ATCCGCTGA CGCTGTCCAC
 501 TTCAAACCTG CCAAGCCTT TTCAATCAA TGCATTGACT TCTCAAACT

551 GGCATTGGGA TTCGGGTTGG AACCTCTAA ACATCATCGG GAACAAATAA

This encodes a protein having amino acid sequence <SEQ ID 370>:

5 1 MAFITRLFKS IKQWLVLPLM LSVLPDAAAE GIDVSRAEAR IXDGGQLSXX
 51 SRFQTELPDQ LQAXXRGVX LNKTLXWQLS APTIASYRFK LGQLIGDDDX
 101 IDYKLSFHPL TNRYRVTGVA FSTXYDTLDA ALRATGAVAN WKVLNKGALS
 151 GAEAGETKAE IRLTLSSTKL PKPFQINALT SQNWHLDSGW KPLNIIGNK*

Homology with a predicted ORF from *N.gonorrhoeae*

ORF106 shows 90.5% identity over a 199aa overlap with a predicted ORF (ORF106.ng) from *N.*

10 *gonorrhoeae*:

 orf106.pep MAFITRLFKSSK-WLIVPLMLPAFQNVAAEGIDVSRAEARITDGGQLSISSRFQTELPDQ 59
 ||||||| | |:: :| :: :||||| :||||||| :||||||| :|||||||
 orf106.ng MAFITRLFKSIKWLVLLPILSVLPDAAAEGIAATRAEARITDGGRLSISSRFQTELPDQ 60
 orf106.pep LQQALRRGVPLNFTLSWQLSAPTIASRYRFKLGQLIGDDDNIDYKLSFHPLTKRXYRVTGVA 119
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 orf106.ng LQQALRRGVPLNFTLSWQLSAPTIASRYRFKLGQLIGDDDNIDYKLSFHPLTKRXYRVTGVA 120
 orf106.pep FSTDYDTLDAALRATGAVANWKVLNKGALSGAEAGETKAEIRLTLSSTKLKPKPFQINALT 179
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 orf106.ng FSTDYDTLDAALRATGAVANWKVLNKGALSGAEAGETKAEIRLTLSSTKLKPKPFQINALT 180
 orf106.pep SQNWHLDSGWKPLNIIGNK 198
 ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 orf106.ng SQNWHLDSGWKPLNIIGNK 199

Due to the K→N substitution at residue 111, the homology between ORF106ng and ORF106-1 is 91.0% over the same 199 aa overlap.

The complete length ORF106ng nucleotide sequence <SEQ ID 371> is:

30 1 ATGGCTTTTA TTACGCGCTT ATTCAAAAGC ATTAAACAAT GGCTTGTGCT
 51 GTTGCCGATA CTCTCCGTTT TGCCGGGACG GGCGGGCGAG GGCATTGCCG
 101 CGACCCGCGC CGAAGCGAGG ATAACGACG CGCGGGCGGT TTCCATCAGC
 151 AGCGCGTTC AAACCGAGCT GCCCGACGAG CTCCAACAGG CGTTCGCGCG
 201 GGCGGTCACG CTCACTTTTA CCTTACGCTT CGACGCTTTC GCCCGACACA
 251 TCCGCTCTTA TCGGTTTAA TTGGGCAAC TCGATTGGCA TGACGACAA
 301 ATTGACTACA AAGTAAGTTT CCATCCGCTG ACCAACCGCT ACCGCTTAC
 351 CGTCGCGCA TTTCCACCG ATTACGACAC TTTGGATGCG GCATTGGCGG
 401 CGACCGCGCG GGTTGCCAAC TGAAGAGTCC TGAACAAGAG CGCGTGTGCC
 451 GGTGGCGAAG CAGGGGAAC CAAGGCGGAA ATCCGCTGGA CGCTGTCCAC
 501 TTCAAAACTG CCCAAGCCTT TCCAAATCAA CGCATTGACT TCTCAAAACT
 551 GGCATTTGGA TTCGGGTTGG AAACCTCTAA ACATCATCGG GAACAAATAA

This encodes a protein having amino acid sequence <SEQ ID 372>:

45 1 MAFITRLFKS IKQWLVLPLI LSVLPDAAAE GIAATRAEAR ITDGGRLSIS
 51 SRFQTELPDQ LQQALRRGVP LNFTLSWQLS APTIASYRFK LGQLIGDDDN
 101 IDYKLSFHPL TNRYRVTGVA FSTDYDTLDA ALRATGAVAN WKVLNKGALS
 151 GAEAGETKAE IRLTLSSTKL PKPFQINALT SQNWHLDSGW KPLNIIGNK*

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF106-1 (18kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 13A shows the results of affinity purification of the His-fusion protein, and Figure 13B shows the results of expression of the GST-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for FACS analysis (Figure 13C). These experiments confirm that ORF106-1 is a surface-exposed protein, and that it is a useful immunogen.

Example 44

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 373>:

```

10 1 ATGGACACAA AAGAAATCCT CGG.TACGCG GcAGGcTCGA TCGGCAGCGC
51 GGTTTTAGCC GTCATCATCC TGCCGCTGCT GTCGTGGTAT TTCCCGCGCG
101 ACGACATCGG GCGCATCGTG CTGATGAGCA CGCGCGCGGG GCTgACGGTG
151 TCGGTGTGTG GCTCGGGGCT GGATCAGGCA TACGTCGCGG AATACTATGC
201 CACGCGCGAC AAGACAcCT TGTTCAAAAC CCGTCTCTCTG CGCGCGCTGC
15 251 TGTCTGCGCG CGCGATAGCC GCCCTGCTGC TTTCGCGGCC GTCCCTGCGG
301 TCTGAAATCC TGTTTTCACT CGACGATGCC gCGCGCGGCA TCGGGCTGGT
351 GCTGTTTTGAA CtGAGCTTCC TGCCCATCGG CTTCCTCTTA CTGGTTTTGC
401 GTATGGAAGG ACGGCGCCTT GCCCTTTCGT CGCGCAACT CCGTGCACAG
451 CTGCGCATCC TGCTGCTG. T CGGCTTGAAG GTGCGGCTGC TGCACTTTCC
20 501 ACGGACACCC CGGCTCCTGA CGCGCGTTTA CGCGCTGGCA AACCTTGCCG
551 CGGCGCCTT TTTGCTGTTT CAAACCCGAT GCCGTCTGAA GGCGTCCGG
601 CACGACCCGT TTTGCGCCGC GTCCTGCAC CGGGG. TGC GCTACGGCAT
651 ACCGATCGCA CTGAGCAGCA TCGCCTATTG GGGGCTGGCA TCGCGCGACC
701 GTTTGTTCCT GAAAAAATAT CGCGGCTGG AACAGCTCGG CGTTTATTGG
25 751 ATGGGTATTT CGTTCGGCGG GCGGCGCATTA TGTTCOCAAA GCATCTTTTC
801 AACGGTCTGG ACACCGTATA TTTTCCGCGC AATCGAAGAA AACGCCCGCG
851 CGGCTCGCCT CTGCGCAACG GCAGAAATCG CGCGCGCCTC GCTTGCTCTCC
901 GCCCTCTGC. TGACCGGCAT TTTCTCGCCC CTGCTCTCCC TCCTGCTGCC
951 GGAATAATAC GCCCGCCTCC GGTTTATCGT CGTATCGTGT ATG. TGCGCCG
30 1001 CGCTGTTTTG CACGCTGGGG GAATCAGCG GCATCGGTTT GAACGTCGTT
1051 CGCAAAACGC GCCCGATCGC GTCCTGCACC TTGGGCGCGC TCGCGCAAAA
1101 CCGTCTGCTG CTGGGGCTGG ACCGTGCGGT CGCGGGGAGG CGCG. GCGG
1151 CGCGGTGTCG CTAATGCGCG TGTTTTTGCG TGTTCAGACC TCTCAGACC
1201 GAAGCTCCTT GCGGCTGTG GCAGCGCTC AAACGCGCTG CGCTTTATCT
35 1251 GCACACATTG TTTCTGCTGA CCGCTCGCG GCGCTACACC TCGTTGCGGA
1301 CGCGGCAAAA CTATCCCGTG TTTGCGCGCG TATGGCGCG ATATCTGGCA
1351 GGCTGCATCC TGCGCCACCG GAAGATTG CACAAACTGT TTCATATTAT
1401 GAAAAACAA GGTTCCTCAT TATGA

```

This corresponds to the amino acid sequence <SEQ ID 374; ORF10>:

```

40 1 MDTKEILXYA AGSIGSAVLA VIILPLLSWY FPADDIGRIV LMQTAAGLTV
51 SVLCLGLDQA YREYYATAD KDTLFKTLFL PPLLSAAALA ALLSRPFLSP
101 SEILFLSDDA AAGIGLVLEF LSLFLPFL LVLRMGRAL AFSSAQLVPE
151 LAIILLXLPT VGLLHFPANT AVLTAVALA NLAAAAFLLF QNRCLRKA
45 201 HAPFSPAVLH RGXRYGPIA LSSIAYWGLA SADRLLFKKY AGLEQLGVYS
251 MGISFGGAAL LFQSIPTVW TPIYFRATEE NAPPARLSAT AESAALLAS
301 ALCXTGIFSP LASLLLPENY RAVRFIVVSC MKPFLCTLA EISGIGLNVV
351 RKRTRPIALAT LGALRANLLL LGLDRAVFA PKGAANACNA SEWLFNAFT
401 ESSCRLWOPL KRLPLVLTFL FCLTSSAAYT CPGTPANYFL FPGVWAAYLA
451 GCILRRRDL HLFHYLKKQ GPPL*

```

Further sequence analysis revealed the complete DNA sequence <SEQ ID 375> to be:

```

1 ATGGACACAA AAGAAATCCT CGGCTACGCG GCAGGCTCGA TCGGCAGCGC
51 GGTTTTAGCC GTCATCATCC TGCCGCTGCT GTCGTGGTAT TTCCCGCGCG
101 ACGACATCGG GCGCATCGTG CTGATGAGCA CGCGCGCGGG GCTGACGGTG

```

```

151 TCGGTGTTGT GCCTCGGGCT GGATCAGGCA TACGTCCGCG AATACTATGC
201 CACCGCCGAC AAGACACCTT TGTTCAAAC CCGTGTCTGC CCGCGCGTGC
251 TGTCTGCCGC CGGATAGTCC GCGCTGCTGC TTTCCGCGCC GTCCCTCGCG
301 TCTGAATACC TGTTTTCACT GCACGATGCC GCGCGCGGCA TCGGGCTGGT
5 351 GCTGTTTGAA CTGAGCTTCC TGCCCATCGG CTTTCTCTTA CTGTTTGTGC
401 GTATGGAAGC ACGCGCCCTT GCCTTTTCGT CCGGCGAACT CGTGCCCAAG
451 CTCGCCATCC TGCTGCTGCT GCGCTGAAC GTCGGCTGCG TGCACTTTCC
501 AGCGAACACC GCGCTCTGA CCGCGTTTA GCGCTGGCA AACCTTGCGG
10 551 CCGCGCGCTT TTTGCTGTT CAAAACCGAT GCGCTGTA GCGCGTCGG
601 CACGACCGCT TTTGCGCCGC CTCTCTGCAC CGGGGCTGC GCTACGGCAT
651 ACGATCGCA CTGAGCAGCA TCGGCTATTG GGGGCTGGCA TCGCGGACCC
701 GTTTGTTCCT GAAAAAATAT GCGCGCTGG AACAGCTCGG CGTTATTATG
751 ATGGGTATTG OGTTCCGGCG GCGGCGATTA TTGTTCCAAA GCATCTTTTC
15 801 AAGGTTCTGG ACACCGTATA TTTTCCGCGC AATCGAGAA AACCGCCGCG
851 CCGCGCCGCT CTGGCGAACC GCAGATCCG CCGCGCGCTT TTTGCTGCTC
901 CGCCTCTGCC TGACCGGAT TTTCTCGGCC CTGCTCTCCC CTCTGCTGCC
951 GGAATACTAC CTGGCGGAT GGTTTATCGT CGTATCGTGT ATGCTGCGCG
1001 CGCTGTTTTC CACGCTGGCG GAAATCAGCG GCATCGTGT GAACGTCGTC
20 1051 CGCAAAACGC GCGCGATCGC GCTCGCCACC TTGGGGCGCG CGCGGCGCGG
1101 CTGCTGCTGC CTGGGCGCTG CCGTGCCGTC CCGCGCGCGG CGCGGCGCGG
1151 CGGTGCGCTG TGCGGCCCTA TTCTGGCTGT TTTTGCCTT CAAGACCGAA
1201 AGCTCCTGCC GCGCTGGGCA GCGGCTCAA CGCTCGCGC TTTATCTGCA
1251 CACATTGTTC TGCTGACCT CTCTGCGCGC CTACACTGCG TTCGCGAGC
1301 CGGCAAACTA TCCCTGTGTT GCGCGCTAT GGGCGGCATA TCTGGCAGCG
25 1351 TGACCTCGC GCCACCGGAA AGATTGTCAC AAACGTGTT ATTATTGAA
1401 AAAACAAGGT TTCCATTAT GA

```

This corresponds to the amino acid sequence <SEQ ID 376; ORF10-1>:

```

1 MDTKEILGYA AGSIGSAVLA VIILPLLSWY FPADDIGRIV LMQTAAGLTV
51 SVLCLGLDQA YVREYYATAD KDTLFTLFL PFLLSAAAIA ALLSRPFLP
30 101 SEILFLSLDA AAGIGLVLEF LSLPIRFL LVLMEGRAL AFSSAQLVFK
151 LAILLLLPLT VGILLHFPANT AVLTAVYALA NLAAALFL QNCRKLAVF
201 HAFPSFAVLH RGLRYGIEIA LSSIAWGLA SADLEITA AGLBQLAVF
251 MGLSFGNAL LFGSIFSTVW TFIYFRAIE NAPPARLST ABSAALLAS
35 301 ALCTGIFSE LASILLPENY AAVRFIVVSC MLPLFCFLA EISGIGLNVV
351 RKRTRPIALAT LGALANLLL LGLAVPSGGA RGAAVCAAS FWLFFAFKT
401 SSRLWQPLK RLPLYLHTLF CLTSSAAYTC FGTANYPLF AGVWAAAYLAG
451 CILRHRKDLH KLFHYLKKQG FPL*

```

Computer analysis of this amino acid sequence gave the following results:

Prediction

ORF10-1 is predicted to be the precursor of an integral membrane protein, since it comprises several (12-13) potential transmembrane segments, and a probable cleavable signal peptide

Homology with EpsM from *Streptococcus thermophilus* (accession number U40830).

ORF10 shows homology with the epsM gene of *S. thermophilus*, which encodes a protein of a size similar to ORF10 and is involved in expolysaccharide synthesis. Other homologies are with prokaryotic membrane proteins:

Identities = (25%)

```

Query: 213 LRYGIPLALSSLAYWGLASADRLFLKKYAGLEQLGVMSIGISFGGAALLLQSFSTVW 270
      L Y +PL SS+ +W L ++ R F+ + G G+ ++ + +IF+ W
Sbjct: 210 LYYALPLIPSSILWLLNASSRYVFLFLLGAGANGLLAVATKPIISIIISINTIFTQAW 267

```

Identities = 15/57 (26%), Positives = 31/57 (54%)

```

Query: 7 LGVAAGSIGSAVLAVIILPLLSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQAYVR 63
      L Y +G+GS +L +++PL ++ + G L QT A L + + + + +R
Sbjct: 12 LVFTIGNLGSKILVFLVFLVLTXYMTQPEYGMADLYQTANLLPLITMNVFADATLR 68

```

Identities = 16/96 (16%), Positives = 36/96 (37%)

Query: 307 IFSPLASLLPENYAARFVVSVCMLPPLFYTLTEISGIGLNVVRKTRPIXXXXXXXXX 366
 + P+ ++ +YA+ V ML LF + ++ G ++T+ +
 Sbjct: 305 VLKPIVEKVSSDYASSWQYVFFMLSMLFSFSDFFGTNYIAAKQTKGVFMTSIYGTIV 364

Homology with a predicted ORF from *N.meningitidis* (strain A)ORF10 shows 95.4% identity over a 475aa overlap with an ORF (ORF10a) from strain A of *N.**meningitidis*:

		10	20	30	40	50	60
orf10.pep		MDTKEILXYAAGSIGSAVLAVIILP	LLSWYFPADDIGRIVIMQTAAGLT	VSVLCLGLDQA			
orf10a		MDTKEILGYAAGSIGSAVLAVITL	PLLSWYFPADDIGRIVLMQTAAGLT	VSVLCLGLDQA			
		10	20	30	40	50	60
orf10.pep		YVREYYATADKDTL	FKTLFLPPLLSAAAI	AALLSRP	SLPSEILF	SLDDAAAGIG	LVLF
orf10a		YVREYYAADKDTL	FKTLFLPPLLSAAAI	AALLSRP	SLPSEILF	SLDDAAAGIG	LVLF
		70	80	90	100	110	120
orf10.pep		LSFLPIRIFLLVL	RMEGRALAFSSAQL	VPSKAILLLXPL	TVGLLHFP	PANTAVLT	AVAYALA
orf10a		LSFLPIRIFLLVL	RMEGRALAFSSAQL	VPSKAILLLXPL	TVGLLHFP	PANTAVLT	AVAYALA
		130	140	150	160	170	180
orf10.pep		NLAAAAFLFQNR	CRILKAVRHAPF	SPAVLHRGX	RGIPIALSS	IYWG	LASADRLFLK
orf10a		NLAAAAFLFQNR	CRILKAVRHAPF	SPAVLHRGX	RGIPIALSS	IYWG	LASADRLFLK
		190	200	210	220	230	240
orf10.pep		AGLEQLGVYS	MGISFGGAALL	FQSIFSTV	WTVPYIFRA	IEANAP	PARLSATAESAA
orf10a		AGLEQLGVYS	MGISFGGAALL	FQSIFSTV	WTVPYIFRA	IEANAP	PARLSATAESAA
		250	260	270	280	290	300
orf10.pep		ALCXTGIFSP	PLASLLPENYA	ARFIVVSC	MPLECTLAE	ISGIGLNV	VVRKTRPIAL
orf10a		ALCXTGIFSP	PLASLLPENYA	ARFIVVSC	MPLECTLAE	ISGIGLNV	VVRKTRPIAL
		310	320	330	340	350	360
orf10.pep		LGALAANLL	LLGLDRAV	PAR-PXGA	AAVACAAS	FWLFFA	FKTESCR
orf10a		LGALAANLL	LLGLDRAV	PAR-PXGA	AAVACAAS	FWLFFA	FKTESCR
		370	380	390	400	410	419
orf10.pep		LFCLTSSA	AYTCFGT	PANYPLF	AGVWAAY	LASCIL	RHRKDLHKL
orf10a		LFCLTSSA	AYTCFGT	PANYPLF	AGVWAAY	LASCIL	RHRKDLHKL
		420	430	440	450	460	470
orf10.pep		LFCLTSSA	AYTCFGT	PANYPLF	AGVWAAY	LASCIL	RHRKDLHKL
orf10a		LFCLTSSA	AYTCFGT	PANYPLF	AGVWAAY	LASCIL	RHRKDLHKL

The complete length ORF10a nucleotide sequence <SEQ ID 377> is:

1 ATGGACACAA AGAAATCCT CGGTACGCG GCGAGTCTCGA TCGGCAGCGC
 51 GGTTTTATGCC GTCATCATCC TGCCGCTGCT GTCCTGGTAT TTCCCTGCGC
 101 ACGACATCGG ACGCATCGTG CTGATGCRGA CGGCGCGCGG GCTGACGCGG
 151 TCGGTGTGTG GCCTCGGGCT GGATCAGGCA TACGTCGCGC AATACTATGC
 201 CGCCGCGGAC AAAGACACTT TGTTCAAAAC CCTGTCTCTG CGCGCGCTGC
 251 TGCTCTGCCG CGGATAGGCC GCCCTGCTGC TTTCCGCCCC ATCCCTGC

301 TCTGAATATCC TGTTTTCGCT CGACGATGCC GCGGCCGSCA TCGGGCTGGT
 351 GCTGTTTGAAC CTGAGCTTCC TGCCCATCCG CTTTCTCTTA CTGGTTTTGC
 401 GTATGGAAGC AGCGGCTTCT GCGCTTTCGT CCGCGCAACT CGTGTCCAAG
 451 CTGCGCATCC TGCTGCTGCT GCGCGTGAOG GTGCGGCTGC TGCACTTTCC
 501 GCGCAACACC GCGCTCTGA CGCGCTGGCA AACCTTCCGG
 551 CCGCGGCTTT TTTGCTGTTT CAAACCGAT GCCGTCTGAA GGCCGTCCGG
 601 CGCGCACCGT TTTCTCCGCT CGTCTGCGAT CGCGGCTGCG GTACAGGAT
 651 ACCGATCGCA CTAAGCAGCA TCGCTTATTT GGGGCTGGCA TCCGCGCACC
 701 GTTGTGTTCT GAAAAATAT GCGGCGCTAG AACAGCTCGG CGTTTATTCTG
 751 ATGGGTATTT CGTTCGGCGG AGCGGCATTA TTGTTCCAAA GCATCTTTTCT
 801 AACGCTCTGG ACACCGTATA TTTTCCGCGC AATCGAAGCA AACGCCCTCG
 851 CCGCGCGGCT CTCGGCAACG GCAGAAATCC CCGCGCGGCT GCTTGCCTCC
 901 GCGCTCTGCC TGACCGGCAT TTTCTCGGCC CTGCGCTCCC TCCTGCTGCC
 951 GGAATACTAC GCGCGCGTCC GGTTTATCGT CGTATCGTGT ATGCTGCCTC
 1001 CGCTGTTTGT CAGCTGGTGA GAAATCAGCG GCATCGGTTT GAAGCTGCTC
 1051 CGAAAAACAC GCGCGATCGC GCTCGCCACC TTGGCGCGCG TGCGCGCAAA
 1101 CCGTCTGCTG CTGGGGCTTG CCGTACCGCT CGCGCGCGCG CGCGCGCGCG
 1151 CGGTTGCGCT TGCGGCTCTA TTTTGGCTGT TTTTGTGTTT CAGACCGGCG
 1201 AGCTCTCGCC TGCGTGGGCT CCGCGCGGCG CTACACCTGC TTGCGCATC
 1251 CACATTGTTC TGCGTGGGCT GCGCGCGTAT GCGCGGTATA TCTGGCAGCG
 1301 CGGCAACTC CCGCGGTGTT GCGCGGTGAT GCGCGGTATA TCTGGCAGCG
 1351 TGCACTCTGC GGCACCGGAA AGATTGTCAC AAATCTTTTC ATTATTGGA
 1401 AAAACAAGT TTCCCATAT GA

This encodes a protein having amino acid sequence <SEQ ID 378>:

1 MDTKEILGYA AGSIGSAVLA VIILPLLSWY FPADDIGRIV LMQTAAGLTV
 51 SVLCGLGDOA YVREYYAAAD KDTLFTKLFL PPLLSAAIA ALLSRPSLP
 101 SEILFSLDDA AAGIGLVLEF LSLFLPIRFL LVLRMEGRAL AFSSAQLVSK
 151 LAILLLELFT VGLLHFPANT AVLTAVYALA NLAAAFLFL QNRCLKAVR
 201 RAPFSSAVLH RGLRYGPIA LSSIAYWGLA SADRLFLKXY AGLEQLGVYS
 251 MGISFGGAAL LFQSFSTVW PYIFRAIEA NAPPARLSAT AESAAALLAS
 301 ALCLTGIFSP LASLLPENY AAVRFIVVSC MLPELCTLV EISGIGLVNV
 351 RKTRPIALAT LGALAAANLL LGLAIVPSGGA RGAAVACAAS FWLFPVPTKE
 401 SSRLWQQLK RLFLYMHFTL CLASSAAATC FGTANYPLF AGVWAVYLAG
 451 CILHRKDLH KLFHYLKKQG FFL*

ORF10a and ORF10-1 show 95.4% identity in 475 aa overlap:

	10	20	30	40	50	60
orf10-1.pep	MDTKEILXYAAGSIGSAVLA	VIILPLLSWYFPADDIGRIV	LMQTAAGLTVSVLCGLGDOA			
orf10a	MDTKEILGYAAGSIGSAVLA	VIILPLLSWYFPADDIGRIV	LMQTAAGLTVSVLCGLGDOA			
	10	20	30	40	50	60
orf10-1.pep	YVREYYATADKDTLFTKL	FLPPLLSAAIAALLSRPSLP	SEILFSLDDAAGIGLVLEF			
orf10a	YVREYYAAADKDTLFTKL	FLPPLLSAAIAALLSRPSLP	SEILFSLDDAAGIGLVLEF			
	70	80	90	100	110	120
orf10-1.pep	LSFLPIRFLLLVLRMEGRAL	AFSSAQLVSKLAILLXPLTV	GLLHFPANTAVLTAVYALA			
orf10a	LSFLPIRFLLLVLRMEGRAL	AFSSAQLVSKLAILLXPLTV	GLLHFPANTAVLTAVYALA			
	130	140	150	160	170	180
orf10-1.pep	NLAAAFLFLFQNRCLKAVR	HAPFSSAVLHRGXRYGIP	IALSSIAIWGLASADRLFLKXY			
orf10a	NLAAAFLFLFQNRCLKAVR	HAPFSSAVLHRGXRYGIP	IALSSIAIWGLASADRLFLKXY			
	190	200	210	220	230	240
orf10-1.pep	AGLEQLGVYSMGISFGGAAL	LFQSFSTVWPYIFRAIEAN	APPARLSATAESAAALLAS			
orf10a	AGLEQLGVYSMGISFGGAAL	LFQSFSTVWPYIFRAIEAN	APPARLSATAESAAALLAS			
	250	260	270	280	290	300
orf10-1.pep						
orf10a						
	310	320	330	340	350	360

5	orf10-1.pep	ALCXTGIFSPLASLLPENYA	AVRFIVVSCMXPLP	FCTLAEISGIGLNVVRKTRP	IATAL	
	orf10a	ALCLTGIFSPLASLLPENYA	AVRFIVVSCMLP	FLFCTLVEISGIGLNVVRKTRP	IATAL	
10	orf10-1.pep	LGALAAANLLLLGLDRAV	PAR--PXGA	AVACAASF	WLF	FAFKTESSCRLWQPLKRLPLYLHT
	orf10a	LGALAAANLLLLGL--AV	PSGGARGA	AVACAASF	WLF	FFVKTESSCRLWQPLKRLPLYMHT
15	orf10-1.pep	LFCLTSSAAYTCFGT	PANYPLFAGV	WAAYLAGC	ILHRKDLHKL	FHYLKKQGFPLX
	orf10a	LFCLASSAAYTCFGT	PANYPLFAGV	WAAYLAGC	ILHRKDLHKL	FHYLKKQGFPLX

Homology with a predicted ORF from *N.gonorrhoeae*

ORF10 shows 94.1% identity over a 475aa overlap with a predicted ORF (ORF10.ng) from *N.*

20	gonorrhoeae:					
	orf10ng.pep	MDTKEILGYAAGSIGSAVLAVIIL	PELLSWYFPADDIGRIVM	QTAAGLTVSVLCLGLDQA	60	
25	orf10nm	MDTKEILXYAAGSIGSAVLAVIIL	PELLSWYFPADDIGRIVM	QTAAGLTVSVLCLGLDQA	60	
	orf10ng.pep	YVREYYAAADKDTLFKTLF	LPPLLFSAAIAALLSRPS	LPSLSEILFSLDDAAAGIGLV	120	
30	orf10nm	YVREYYATADKDTLFKTLF	LPPLLFSAAIAALLSRPS	LPSLSEILFSLDDAAAGIGLV	120	
	orf10ng.pep	LSFLPIRIFLLLVLRMEGRALAFSSA	QLVPKLAILLPLTVGLLHF	PANTSVLTAVYALA	180	
35	orf10nm	LSFLPIRIFLLLVLRMEGRALAFSSA	QLVPKLAILLPLTVGLLHF	PANTSVLTAVYALA	180	
	orf10ng.pep	NLAAAFLFQNRCLKAVRRAPFS	PAVLHRGLRYGIPALSL	SAYWGLASADRLFLKKY	240	
40	orf10nm	NLAAAFLFQNRCLKAVRHAFPS	PAVLHRGXRYGIPALSL	SIAYWGLASADRLFLKKY	240	
	orf10ng.pep	AGLEQLGVYSMGISFGGAALLQS	IFSTVMTPIYIFRAIENAT	PARLSATAESAAALLAS	300	
45	orf10nm	AGLEQLGVYSMGISFGGAALLQS	IFSTVMTPIYIFRAIENAT	PARLSATAESAAALLAS	300	
	orf10ng.pep	ALCLTGIFSPLASLLPENYA	AVRFTVVS	CMLPFLFYTLTEISGIGLNVVRKTRP	IATAL	360
50	orf10nm	ALCXTGIFSPLASLLPENYA	AVRFIVVSCMXPLP	FCTLAEISGIGLNVVRKTRP	IATAL	360
	orf10ng.pep	LGALAAANLLLLGL--AV	PSGGTRGA	AVACAASF	WLF	FAFKTESSCRLWQPLKRLPLYLHT
55	orf10nm	LGALAAANLLLLGLDRAV	PAR--PXGA	AVACAASF	WLF	FAFKTESSCRLWQPLKRLPLYLHT
	orf10ng.pep	LFCLASSAAYTCFGT	PANYPLFAGV	WAAYLAGC	ILHRKDLHKL	FHYLKKQGFPLX
60	orf10nm	LFCLTSSAAYTCFGT	PANYPLFAGV	WAAYLAGC	ILHRKDLHKL	FHYLKKQGFPLX

The complete length ORF10ng nucleotide sequence <SEQ ID 379> is:

60	1	ATGGACACAA	AAGAAATCCT	CGCTACGCG	GCAGGCTCGA	TGGCGAGCGC
	51	GGTTTTAGCC	GTCATCATCC	TCCGCTGCT	GTCGTGGTAT	TTCCCGCGCG
65	101	ACGACATCGG	GCGCATCTGT	CTGATCGAGA	CGCGGCGGCG	ACTGACGGTG
	151	TCGGTATGTG	GCCTCGGGCT	GGATCAGGCA	TACGTCGGCG	ATACTATGCG
70	201	CGCCGCGCAG	AAAGACACTT	TGTTCAAAC	CCTGTCCTCG	CGCGCGCTCG
	251	TGTTTTCCGC	CGCGATAGCC	GCCCTGCTCG	TTTCCGCCCC	GTCCCTCGCG
75	301	CTGAAATACC	TGTTTTGCGT	CGACGATGCC	GCCGCGCGCA	TCGGGCTGGT
	351	GCTGTTTGA	CTGAGCTTCC	TGCCCATCCG	CTTTCCTTA	CTGTTTTTCG

401 GTATGGAAGG GCGGCGCCTT GCCTTTTCGT CGCGGCAACT CGTGCCCAAA
 451 CTCGCCAATC TGCTGCTGTT GCGGCTGACG CTGCGGGTGC TGCACCTTTC
 501 GCGGAACACC TCGCTCTGTA CGCGCCTTTA CGCGCTGCGA AACCTTGCGG
 551 CGCGCCGCTT TTTGCTGTTT CAAACCCGAT GCGCTCTGAA GCGCGTCGG
 601 CGCGGCGCCT TTTGCGCCCG CTGCTGTCAC CGGCGGTGC GCTACGCAT
 651 ACCGCTCGCA CTGAGCAGCC TTGCTATTGT GGGGCTGCGA TCCCGCGACC
 701 GTTTGTTCCT GAAAAAATAT GCGGCGCTGG AACAGCTCGG CGTTATTTCG
 751 ATGGGTATTT CGTTCGCGGG GCGGCGCATTA TTGCTCCAAA GCATCTTTTC
 801 AACGGTCTGG ACACCGTATA TTTTCCGTGC AATCGAAGAA AACGCCAGCG
 851 CGCGCCGCGT CTGCGGCAAG CGAGAATCCG CGCGCGCCTT GCTTGCTTCC
 901 GCGCTCTGCG TGACCGGAAT TTTCTGCGCG CTGCGCTCCC TCCTGCTGCC
 951 GGAAAACTAC GCGCGCGCTC GGTPTACCGT GATATCGTGT ATGCTGCGCG
 1001 cgctGTTTTA CACGCTGACC GAAATCAGCG GCATCGGTTT GAACGCTGTC
 1051 CGCAAAACGC GTCCGATCGC GCTTGCCACC TTGGGCGCGC TGGCGGCAAA
 1101 CCTGCTGCTG CTGGGGCTTG CGGTACCGTC CGGCGCGCAC GCGCGCGCGG
 1151 CGGTGCGCTG TGCGCGCTCA TTCTGTTGT TTTTGTGTTT CAAGACAGAA
 1201 AGCTCCTGCC GCCTGTGGCA CGCGCTCAAA CGCCTGCGCG TTTATATGCA
 1251 CACATTTGTC TGCCTGCGCT CTGCGCGCGC CTACACCTGC TTGCGCACAC
 1301 CGCAAACTA CCCTGCTT gctggcgctat gctggcgctat TCTGCGAGCG
 1351 TGCATCTGCG GCCACCGGAA AAATTGCGAC AAACCTGTTT ATTATTGAA
 1401 AAAACAAGT TTCCATTAT GA

This encodes a protein having amino acid sequence <SEQ ID 380>:

1 MDTKEILGYA AGSIGSAVIA VIILPLLSWY FPADDIGRIV IMQTAAGLTV
 5 SVLCLGLDQA VYREYYAAD KDTLFTLFL PELLFSAAIA ALLSRPSLSP
 10 SEILFSLDDA AAGIGLVFE LSFLPIRFL ILVLRMEGRAL AFSSAQLVPK
 15 LAILLPLLT VGLLHFPANT SVLTAVYALA NLAAAFLLF QNRCLRAVR
 20 RAPFSPAVLH RGLRYGIPLA LSSLAYWGLA SADRLFLKKY AGLEQLGVYS
 25 MGISFGGAAL LQISFSTVW TFIYFRAIEE NATPARLSAT AESAAAILAS
 30 ALCLTGIFSP LASLLPENY AAVRFTVVS CMLPFLYTLT EISGIGLVNV
 35 RTRTPALAT LGALAANLLL LGLAVPSSGT RGAACACAS FWLFVFKE
 40 SSCLRWQLK RLPLYMHTL CLASSAYTC FGTFANYPLF AGVWAAYLAG
 45 CILHRKRLNH KLHYHLKQG FFL*

ORF10ng and ORF10-1 show 96.4% identity in 473 aa overlap:

35	orf10-1.pep	MDTKEILGYAAGSIGSAVIAVIILPLLSWYFPADDIGRIVIMQTAAGLTVSVLCLGLDQA	10	20	30	40	50	60
	orf10ng-1	MDTKEILGYAAGSIGSAVIAVIILPLLSWYFPADDIGRIVIMQTAAGLTVSVLCLGLDQA	10	20	30	40	50	60
40	orf10-1.pep	VYREYYATADKDTLFTLFLPPLLSAAIAALLSRPSLSEILFSLDDAAAGIGLVFE	70	80	90	100	110	120
	orf10ng-1	VYREYYAADKDTLFTLFLPPLLSAAIAALLSRPSLSEILFSLDDAAAGIGLVFE	70	80	90	100	110	120
45	orf10-1.pep	LSFLPIRFLLLVLRMEGRALAFSSAQLVPKLAILLPLLTVGLLHFPANTAVLTAVYALA	130	140	150	160	170	180
	orf10ng-1	LSFLPIRFLLLVLRMEGRALAFSSAQLVPKLAILLPLLTVGLLHFPANTSVLTAVYALA	130	140	150	160	170	180
50	orf10-1.pep	NLAAAFLFQNRCLRAVRRAFPSPAVLHRLRGYIPLASSIAYWGLASADRLFLKKY	190	200	210	220	230	240
	orf10ng-1	NLAAAFLFQNRCLRAVRRAFPSPAVLHRLRGYIPLASSLAYWGLASADRLFLKKY	190	200	210	220	230	240
55	orf10-1.pep	AGLEQLGVYMGISFGGAALLFQSIFSTVWTPYIFRAIENATPARLSATAESAAAILAS	250	260	270	280	290	300
	orf10ng-1	AGLEQLGVYMGISFGGAALLFQSIFSTVWTPYIFRAIENATPARLSATAESAAAILAS	250	260	270	280	290	300
60	orf10-1.pep	ALCLTGIFSPLASLLPENYAAVRFTVVS CMLPFLFCTLAISGIGLVNVKTRTPALAT	310	320	330	340	350	360
	orf10ng-1	ALCLTGIFSPLASLLPENYAAVRFTVVS CMLPFLFCTLAISGIGLVNVKTRTPALAT	310	320	330	340	350	360

5	orf10ng-1	ALCLTGTGIFSPLASLLLPENYAARFTVVSCMLPFLFYTLTEISGIGLVNRKTRIPALAT	310	320	330	340	350	360
	orf10-1.pep	LGALAAANLILLGLAVPSGGARGAACAASFWLFFAFKTESSCRWLQPKRLPLYLHTLF	370	380	390	400	410	420
	orf10ng-1	LGALAAANLILLGLAVPSGGTRGAACAASFWLFFVKTESSCRWLQPKRLPLYNHTLF	370	380	390	400	410	420
10	orf10-1.pep	CLTSSAAYTCFGTPANYPLFAGVWAAYLAGCILLRHKDLHLKLFHYLXKQGFPLX	430	440	450	460	470	
	orf10ng-1	CLASSAAYTCFGTPANYPLFAGVWAAYLAGCILLRHKDLHLKLFHYLXKQGFPLX	430	440	450	460	470	

Based on this analysis, including the presence of a putative leader peptide and several transmembrane segments and the presence of a leucine-zipper motif (4 Leu residues spaced by 6 aa, shown in bold), it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 45

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 381>:

25	1..ATCCTGAAAC	CGCATACCA	GCTTAAGGAA	GACATCCAAC	CTGATCCGGC
	51	CGATCAAAAC	GCCTTGTCCG	AACCGGATGC	TGCGACAGAG
	101	CGGATCCGGA	AAATGCTGCC	GACAAGCAGC	CCGTTGCCGA
30	151	GAGGTTGAAG	AAAAGGCGGG	CGAGCCGGAA	CGGAAGAGC
	201	GGCAGTGGCT	AAGAAAGCGC	TGACGGAAGA	GCGTGAACAA
	251	AAAAAGCGCA	GAAGAAAGAT	GCCGAAACGG	TTAAATACAA
35	301	CCGTCTAAGG	AAACAGAGAA	AAAAGCTTCA	AAAGAAAGGA
	351	GAAGGAAAAA	GTTGCACCCA	AACCAACCCC	GGACAAATAC
	401	GCAGCATCGA	AAAmGGCGCG	AgTGCCCGCG	CCAAAGAAATG
40	451	AA..AACCTCC	GACAAAGGCGG	AAGC..AACGC	ATTATCTGCA
	501	TATGCCGACC	GTACAGAGCGT	GCAAGGGGAG	CCTGCCCAAC
	551	GGGCATATCT	TCCAAGGTGG	TCGGTATACA	GGCGGACAT
35	601	ACCGGTGCA	AAGCGCAAT	ATGCTCGCGC	ATCGCGTGA

This corresponds to the amino acid sequence <SEQ ID 382; ORF65>:

40	1..ILKPHNQLKE	DIQPDFADQN	ALSEPDAATE	AEQSDAENAA	DKQPVADKAD
	51	EVEEKAGEPE	REEPDGQAVR	KKALTEEREQ	TVREKAQKQD
	101	PSKETEEKKAS	KEEKKAKEK	VAPKPTPEQI	LNSGSTEXAR
40	151	XNVRQGSXSR	IICKWARMPT	VRARKGSVNP	WQSWAYLPRW
	201	TGCKAAICLP	MR*		SVIRRDIKRF

Further work revealed the complete nucleotide sequence <SEQ ID 383>:

45	1	ATGTTTATGA	ACAAATTTTC	CCAATCCGGA	AAAGGTCGT
	51	CTTCGGTTTG	ATACTGGCGGA	CGGTCAATTAT	TGCCGGTATT
	101	TGAAACCGAG	CGGTCAAAAT	GGTTCACAAA	TCCCGGCTTC
50	151	CCTCGAGAA	CGGAATCCT	GAACCGGAA	AGCCAGCCTA
	201	CCAACCTGAA	CCGGCCGATC	AAAACGCTT	CTCCGAGACC
	251	CAGAGGCAGA	CGAGTCGGAT	GCGGAAAAAG	CTCCGCACAA
55	301	GCGGATAAAG	CCGACGAGGT	TGAAGAAAAG	CCGGCGGAGC
	351	AGAGCCGGAC	GGACAGGCG	TGCGTAAGAA	AGCGGTGAGC
	401	AACCAACCGT	CAGGGAARAA	GCGCAGAGAA	AAGATGCCGA
55	451	AAACAGCGG	TAAACCGCT	TAAAGAAACA	GAGAAAAAAG
	501	AGAGAAAAAG	GCGGCGAAGG	AAAAGTTTGC	ACCCAAACCA
	551	AAATCCTCAA	CAGCGGCAGC	ATCGAAAAAG	GCGGCGAGTC
55	601	GAAGTGCAGA	AAATGAAAAC	GTCCGACAAG	GCGGAGACAA
	651	GCAATGGGCG	CGGTATGCGC	ACCGTCAGAG	GCGGGAAGGG

701 AACTGGCAAT CTTGGGCATA TCTTCCAAGG TGGTCGGTTA TCAGGCGGGA
751 CATAAACGC TTTACCGGGT GCAAGCGGC AATATGCTG CCGATGCGGT
801 GAAAAAATG CAGGACGAGT TGAAAAACA TGAAGTCGG AGCCTGATCC
851 GTTCTATCA AAGCAATAA

5 This corresponds to the amino acid sequence <SEQ ID 384; ORF65-1>:

	1	MFNMKFFQSG	KGLSGPDTFG	ILATVILIAGI	LEYLNSQSGN	AFKIPASSKQ
	51	PAETEILIKPV	NQPKDETFQE	PADQNALSEP	DAATEAEQSD	AEKAAQKQPV
	101	AOKDAEVEEK	AGEPEREEDP	QOAVRKALKP	EEREQTVREK	AQCKKDAETV
	151	QKAVKPSKST	EKKASKEEKK	AAEKVKAPKP	TPEQTLNSGS	TEKARSAAAK
10	201	EVQKMKTSOK	AEATHVYLMK	OYDLRQSAEG	QKAKLAILGI	SSKVVGYQAG
	251	HKTLYRQSG	NMSDADVKMG	ADQLSKHEVA	SLIRISIESK*	

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF65 shows 92.0% identity over a 150aa overlap with an ORF (ORF65a) from strain A of *N.*

15 *meningitidis*:

[illegible]

The complete length ORF65a nucleotide sequence <SEO ID 385> is:

40	1	ATGTTTATGA	ACAAATTTTC	CCAATCCGGA	AAAGTCTGCT	CCGTTTTTTT
	51	CTTCGGTTTG	ATATCTGGGA	CGGTCAATT	TGCGGGTTAT	TGTTTTTATCT
	101	TGAACTGAA	GGTCAAAAT	CGGTCATCAA	TGCGGGTTTCC	CTCGAAGAC
	151	CTTCGAGAA	CGGAATCTCT	GAATACGCTA	ATCCAGCCTA	AGGAAGACAT
	201	CCAACCTGAA	CGGCGCGAT	AAAACGGTAA	CTCGGAACCG	GATGCTCGGA
45	251	AAGAGCAGA	GCATCTGGAT	CGGGAAGAAAG	CTCGCGGAGT	AGCAGCCGTT
	301	GGCGACAAAT	CGCGACGGT	TGCGGAAAG	GGCGACGAGC	CGGAGCGGGA
	351	AAAGTCGCAC	GCAGCAGCG	TCGCGAGAA	AGCATCTGCG	GAAGAGCGTT
	401	AACAAACCGT	CGGCAAGAA	AGAGTCGAA	AGAGTCTTAA	AGAGTCTTAA
	451	AAACAAAGCG	TAAACCATTC	TAAAGAAAC	CTTCAAAGAA	CTTCAAAGAA
50	501	AGGAAAAAG	CGCGAGNAG	AAATACCTTC	ACCCAAACCG	ACCCCGCGAAC
	551	AAATCTCTCA	CAGCGCGAGC	TGATGAAAG	CGCGCACTGC	CGCTCTCAAA
	601	GAAATGCGCA	CAATGAAAC	CGCGCAGAA	CGCGCACTGA	CGCATTTACT
	651	GCAATTTGGC	GGTATTCGCG	ACGCGCGAG	CGCGGAAAGG	CAGCTGCCA
	701	AACTGGCAAT	CTTGGGCATA	TCCTTCAAG	TGCTTCGGTTA	TCAGGCGGCA
55	751	CATAAAAACG	TTTACCGGTT	CGAAGCGGCG	ATATGTTCTG	TCAGGCGGTT
	801	GAATAAAATG	CAGCAGCATG	TGAAAAACA	TGAAGTCGCC	AGCTTGCGTT
	851	GTTCATATCG	BAGCAATTA			

This encodes a protein having amino acid sequence <SEO ID 386>.

1 MFNKFQSG KGLSGFFFL ILATVILI LFYINQSGON AFKIPVPSQ
 51 PAETELKPK NQPKEDIQPE PADQNALSEP DAAKEAQSD AEKAADKQPV
 101 ADKADEVEEK ADEPEREKSD QGAVRKKALT EEREQTVGEK AQKKDAETVK
 151 KQAVKPSKET EKKAKEKKEK AEKEKVAKP TPEQILNNGS IEKARSAAK
 201 EVQKMTPEK AEATHYLQMG AYADRSABG QRAKLALIGI SSKVGVYQAG
 251 HTLYRVQSG NMSADAVKKM QDELKKHEVA SLIRSIKSK*

ORF65a and ORF65-1 show 96.5% identity in 289 aa overlap:

		10	20	30	40	50	60
10	orf65a.pep	MFNKFQSGKGLSGFFFLILATVILIAGILFYINQSGONAFKIPVPSQPAETELKPK					
	orf65-1	MFNKFQSGKGLSGFFFLILATVILIAGILFYINQSGONAFKIPASQPAETELKPK					
		10	20	30	40	50	60
15	orf65a.pep	NQPKEDIQPEPADQNALSEPDAAKEAEQSDAEKAADKQPVADKADEVEEKADEPEREKSD					
	orf65-1	NQPKEDIQPEPADQNALSEPDAATEAEQSDAEKAADKQPVADKADEVEEKAGEPEREED					
		70	80	90	100	110	120
20	orf65a.pep	QGAVRKKALTEEREQTVGEKAQKKDARTVKQAVKPSKETEKKAKEKKAKEKVAKP					
	orf65-1	QGAVRKKALTEEREQTVREKAQKKDARTVKQAVKPSKETEKKAKEKKAKEKVAKP					
		130	140	150	160	170	180
25	orf65a.pep	TPEQILNNGSIEKARSAAAKEVQKMTTPDKAEATHYLQMGAYADRSABGQRAKLALIGI					
	orf65-1	TPEQILNNGSIEKARSAAAKEVQKMTSKAEATHYLQMGAYADRSABGQRAKLALIGI					
		190	200	210	220	230	240
30	orf65a.pep	SSKVGVYQAGHTLYRVQSGNMSADAVKKMQDELKKHEVASLIRSIKSK					
	orf65-1	SSKVGVYQAGHTLYRVQSGNMSADAVKKMQDELKKHEVASLIRSIKSK					
		250	260	270	280	290	

Homology with a predicted ORF from *N.gonorrhoeae*

ORF65 shows 89.6% identity over a 212aa overlap with a predicted ORF (ORF65.ng) from *N. gonorrhoeae*:

		30	40	50	60	70	80
40	ORF65ng	IIAGILLYLNQGGONAFKIPAPSKQPAETELKLNQPKEDIQPEPADQNALSEPDAKE					
	ORF65	ILKPHNQLKEDIQPDPAQNALSEPDAATE					
45		10	20	30			
	ORF65ng	AEQSDAEKAADKQPVADKADEVEEKAGEPEREEDQGAVRKKALTEEREQTVREKAQKKD					
50	ORF65	AEQSDAENAAADKQPVADKADEVEEKAGEPEREEDQGAVRKKALTEEREQTVREKAQKKD					
		40	50	60	70	80	90
55	ORF65ng	AETVKKQAVKPSKETEKKAKEKKAKEKVAKPPTPEQILNNGSIEKARSAAAKEVQKM					
	ORF65	AETVKIQAVKPSKETEKKAKEKKAKEKVAKPPTPEQILNNGSIEKARSAAAKEVQKM					
		100	110	120	130	140	150
60	ORF65ng	KNFGQGGSQRIICKWARMPNGARKGVSVPNWQSWAYLKPWSAIRRDIKFTCAKCAICPP					
	ORF65	XNVQGGGSKRIICKWARMPNVRARKGVSVPNWQSWAYLPRWSVIRRIDIKFTGCKAATCLP					
		160	170	180	190	200	210

ORF65ng MR
 ORF65 MR

An ORF65ng nucleotide sequence <SEQ ID 387> was predicted to encode a protein having amino

5 acid sequence <SEQ ID 388>:

```

1 MFNMKFSQSG KGLSGFFFL ILATVIIAGI LLYLNQGGGN AFKIPAPSKQ
51 PAETELILKLK NQPKEDIQPE PADQNALSEP DVAKEAEQSD AEKAADKQPV
101 ADKADEVEEK AGEPEEREED QGAVRKALTE EEREQTVREK AQKKAETVVK
151 KQAVPKSKET EKKASKEEKK AAEKVAPKP TPEQILNSRS IEKARSAAAK
201 EVQMKMNFQGG GGSQRILCKW ARMPNFGAKG GSVFNWQSWA YLPKWSAIRR
251 DIKRFTACKA AICPEMR*

```

After further analysis, the complete gonococcal DNA sequence <SEQ ID 389> was found to be:

```

1 ATGTTTATGA ACAATTTTC CCAATCCGGA AAGGTCGTGT CCGGTTTCTT
15 51 CTTCCGTTTG ATACTGGCAA CGGTCAATTAT TGCCGGTATT TTGCTTTATC
101 TGAACCAAGG CGGTCAAAAT CGGTTCAAAA TCCCGGCTCC GTCCGAAGCAG
151 CCTGCAGAAA CGGAATCTCT GAAACTGAAA AACCAGCCTA AGGAAGACAT
201 CCAACCTGAA CCGGCGGATC AAAACGCCCT GTCCGAACCGG GATGTTGCGA
251 AAGAGGCAGA GCAGTCGGATC GCGGAAAAAG CTGCCGACAA CGACCCCGTT
301 GCGGACAAAG ccgacagAGT TGAAGAAAG GCGGcgAgc cggAACGgga
351 aGAGCCGGAC ggACAGGCAG TC CGCAAGAA AGCACTGAgc gAAAgcgGTG
401 AACAAACcgt cagggRAAAA CCGCAAGAA AAGATCCGA ACCGgTTAAA
451 AAacaaGcg Caaacccgtc TAAAGAAACA ggaaaaaag cTTcaaaaga
501 aagaaaaaag cgcgcgaaaag aaaAAGTgtc acccaaaaaag accccggaaC
551 aaatcctcaa cagcccgCagc atcgaaaaag cgcgtagtgc cgctgcacaa
601 gaAgTgcAGA AAatgaaaaa ctTggggcaa ggcgGaaGCC aacgcattaT
651 CTGcaaatgg cgcgctatgc cgacgcgtcc gagcgcggaA gggcagcggt
701 ccaaAActgg aAtccttgGgc atatctTccg aagtgtcgG CTATCAGCGG
751 GGACATAAAA CGCTTTACCG CGTGCAAAGC GGCaatatgt ccgcccgtgc
801 gGTGAAAAAA ATGCAGGACG AGTTGAAAAA GCATGGGGTt gcCAGCCTGA
851 TCOGTGcgAT TGAAGGCAAA TAA

```

This encodes the following amino acid sequence <SEQ ID 390>:

```

1 MFNMKFSQSG KGLSGFFFL ILATVIIAGI LLYLNQGGGN AFKIPAPSKQ
51 PAETELILKLK NQPKEDIQPE PADQNALSEP DVAKEAEQSD AEKAADKQPV
101 ADKADEVEEK AGEPEEREED QGAVRKALTE EEREQTVREK AQKKAETVVK
151 KQAVPKSKET EKKASKEEKK AAEKVAPKP TPEQILNSRS IEKARSAAAK
201 EVQMKMNFQGG GGSQRILCKW ARMPNFGAKG GSVFNWQSWA YLPKWSAIRR
251 GHKTLRYVQS GNMSADAVKK MDELKKHGV ASLIRAEIK*

```

ORF65ng-1 and ORF65-1 show 89.0% identity in 290 aa overlap:

```

40 orf65-1.pep      10      20      30      40      50      60
orf65ng-1         MFNMKFSQSGKGLSGFFFLILATVIIAGILLYLNQGGQNAFKIPASSKQPAETELILKPK
|||||
orf65ng-1         MFNMKFSQSGKGLSGFFFLILATVIIAGILLYLNQGGQNAFKIPASSKQPAETELILK
10      20      30      40      50      60

45 orf65-1.pep      70      80      90      100     110     120
orf65ng-1         NQPKEDIQPEPADQNALSEPDAATEAEQSDAEKAADKQPVADKADEVEEKAGEPEEREED
|||||
orf65ng-1         NQPKEDIQPEPADQNALSEPDAATEAEQSDAEKAADKQPVADKADEVEEKAGEPEEREED
70      80      90      100     110     120

50 orf65-1.pep      130     140     150     160     170     180
orf65ng-1         GQAVRKALTEEREQTVREKAQKKAETVVKQAVKPSKETEKKASKEEKKAAAEKVAPKP
|||||
orf65ng-1         GQAVRKALTEEREQTVREKAQKKAETVVKQAVKPSKETEKKASKEEKKAAAEKVAPKP
130     140     150     160     170     180

55 orf65-1.pep      190     200     210     220     230     239
orf65ng-1         TPEQILNSGSIEKARSAAAEVQKMKTSKAEATHYL-QMGAYADRSAGGQRAKLAILG
|||||
orf65ng-1         TPEQILNSRSIEKARSAAAEVQKMKNFQGGGSGRIICKWARMPTVRSAGGQRAKLAILG
60

```

		190	200	210	220	230	240
orf65-1.pep	240	250	260	270	280	290	
5	ISSKVVG	YQAGHK	TLYRV	SGNMS	ADAVK	KKMODE	LKKHEVASLIR
orf65ng-1	ISSEVV	GQAGHK	TLYRV	SGNMS	ADAVK	KKMODE	LKKHGVLRAIEGK
	250	260	270	280	290		

On this basis, including the presence of a putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 46

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 391>:

1	ATGAACCACG	ACATCACTTT	CCTCACCCGT	TTCTACTCG	GTkTCTTGG
51	CGGAACGCAC	TGCATCGSTA	TGTGCGGGCG	ATTAAAGCAGC	GcGTTTga.s
101	TCCAACTCCC	CCGCGATATC	AACCGCTTTT	GGCTGATCCT	GCTGCTTAAC
151	ACAGGACGGG	TAAGCAGTCA	TACGGCAATC	GGCCTGATAC	TCCGATTAA
201	CGGACAGGTC	GGCGTTTCAC	TGAGCAAAAC	CGCGTCTCTG	CAGAAATATT
251	TATACACGGC	CGCCAACTCT	CTGCTGCTCT	TTTTAGGCTT	ATACTTGAGC
301	GGTATTTCCT	CCTTGGCGGC	AAAAATCGAG	AAATCGGCA	AACCGTATG
351	CGGGAACCTG	AACCGGATAC	TCAACCGCTT	GTTACCCATA	AAATCCATAC
401	CGCGCTGCCT	tGCGgtCGGA	ATATTATGGG	CTGGCTGCC	GTGCGGAATG
451	GTTTACACGG	CGTCGCTTTA	CGCGCTGGGA	AGCGGTAGTG	CGGCAACGGG
501	CGGGTTATAT	ATGCTTGCC	TTGCACTGGG	TACGCTGCC	AATCTTTAG
551	CAATCGGCAT	TTTTTCCCTG	CAACTGAAGA	AAATCATGCA	AAACCGATAT
601	ATCCGCTGT	GTACGGGATT	ATCCGTATCA	TTATGGGCAT	TATGGAAACT
651	TGCGCTCTCG	TGGCTGTAA			

This corresponds to the amino acid sequence <SEQ ID 392; ORF103>:

1	MNHDITFLTL	FLLGXFGGTH	CIGMCGGLSS	AFXQQLPPhi	NRFWLILLIN
51	TGRVSSYTAI	GLILGLIGQV	GVSLDQTRVL	QNILYTAANL	LLLFLGLYLS
101	GISSLAARKIE	KIGKPIWRNL	NPIINRLLEPI	KSIPACLAVG	ILWGLPCGL
151	VYSASLYALG	SGSAATGGLY	MLAFALGTLP	NLLAIGIFSL	QLXKIMQNR
201	IRLCTGLSVS	LWALMKLAVL	NL*		

Further work elaborated the DNA sequence <SEQ ID 393> as:

1	ATGAACCACG	ACATCACTTT	CCTCACCCGT	TTCTACTCG	GTtTCTTGG
51	CGGAACGCAC	TGCATCGSTA	TGTGCGGGCG	ATTAAAGCAGC	GCGTTTGGCG
101	TCCAACTCCC	CCGCGATATC	AACCGCTTTT	GGCTGATCCT	GCTGCTTAAC
151	ACAGGACGGG	TAAGCAGTCA	TACGGCAATC	GGCCTGATAC	TCCGATTAA
201	CGGACAGGTC	GGCGTTTCAC	TGAGCAAAAC	CGCGTCTCTG	CAGAAATATT
251	TATACACGGC	CGCCAACTCT	CTGCTGCTCT	TTTTAGGCTT	ATACTTGAGC
301	GGTATTTCCT	CCTTGGCGGC	AAAAATCGAG	AAATCGGCA	AACCGTATG
351	CGGGAACCTG	AACCGGATAC	TCAACCGCTT	GTTACCCATA	AAATCCATAC
401	CGCGCTGCCT	TGCGGTGCGA	ATATTATGGG	GCTGGCTGCC	GTGCGGAATG
451	GTTTACACGG	CGTCGCTTTA	CGCGCTGGGA	AGCGGTAGTG	CGGCAACGGG
501	CGGGTTATAT	ATGCTTGCC	TTGCACTGGG	TACGCTGCC	AATCTTTAG
551	CAATCGGCAT	TTTTTCCCTG	CAACTGAAA	AAATCATGCA	AAACCGATAT
601	ATCCGCTGT	GTACGGGATT	ATCCGTATCA	TTATGGGCAT	TATGGAAACT
651	TGCGCTCTCG	TGGCTGTAA			

This corresponds to the amino acid sequence <SEQ ID 394; ORF103-1>:

1	MNHDITFLTL	FLLGFFGGTH	CIGMCGGLSS	AFALQLPPhi	NRFWLILLIN
51	TGRVSSYTAI	GLILGLIGQV	GVSLDQTRVL	QNILYTAANL	LLLFLGLYLS
101	GISSLAARKIE	KIGKPIWRNL	NPIINRLLEPI	KSIPACLAVG	ILWGLPCGL
151	VYSASLYALG	SGSAATGGLY	MLAFALGTLP	NLLAIGIFSL	QLXKIMQNR

201 IRLCTGLSVS LWALWKLAVL WL*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)ORF103 shows 93.8% identity over a 222aa overlap with an ORF (ORF103a) from strain A of *N.*5 *meningitidis*:

		10	20	30	40	50	60
	orf103.pep	MNH	DITFL	TLFLL	GFGGTH	CIGMCG	LSSAFXXQLP
10	orf103a	MNX	DITFL	TLFLL	GFGGTH	CIGMCG	LSSAFALQLP
		10	20	30	40	50	60
	orf103.pep	70	80	90	100	110	120
		GLIL	GLIGQV	GVSLDQ	TRVLQN	ILYTAAN	LLFLGL
15	orf103a	GLIL	GLIGQV	GVSLDQ	TRVKN	ILYTAAN	LLFLGL
		70	80	90	100	110	120
	orf103.pep	130	140	150	160	170	180
		NPIL	NRLLP	IKSIPAC	LAVGIL	WGWLP	CGLVYS
20	orf103a	NPIL	NRLLP	IKSIPAC	LAVGIL	WGWLP	CGLVYS
		130	140	150	160	170	180
	orf103.pep	190	200	210	220		
		NLLA	IGIFSL	QLXKIM	QNR	YIRLCT	GLSVS
25	orf103a	NLLA	IGIFSL	QLXKIM	QNR	YIRLCT	GLSVS
		190	200	210	220		

The complete length ORF103a nucleotide sequence <SEQ ID 395> is:

30	1	ATGA	ACCANG	ACAT	CACCTT	CCTC	ACCCTG	TTC	CTACTG	GTTT	CTTCGG
	51	CGGA	ACGCAC	TGCA	TCGGTA	TGTG	CGCGCG	ATT	AAGCAG	CGGT	TTGCGC
	101	TCCA	ACTCCC	CCGC	CATATC	AACG	CTTNT	GGCT	GATCCT	GCTG	CTTAAC
	151	ACAG	GACGGG	TAAG	CAGCTA	TACG	GCAATC	GGCT	GATAC	TCGG	ATTAA
	201	CGG	CAGGTC	GGCG	TTTTCAC	TCGA	CCAAAC	CGCG	CTCNTG	CAGA	ATATTT
	251	TATA	CACGGC	CGCA	ACCTC	CTGC	TGCTCT	TTTT	AGGCTT	ATACT	TGAGC
	301	GGT	ATTTCTT	CCTT	GGCGGC	AAAA	ATCGAG	AAAA	ATCGGA	AACG	ATATG
	351	CGCG	AACCTG	AACCG	GATAC	TCA	ACCGCT	GTT	ACCCATA	AAAT	CCATAC
	401	CCGC	CTGCCT	TGCG	GTGCGA	ATAT	TATGGG	GCT	GGCTGCC	GTGC	GGA
	451	GTTT	ACAGCG	CGTC	CGCTTTA	CGCG	CTGGGA	AGCG	GAGTGT	CGGC	AACGGG
	501	CGGG	TTATAT	ATGC	TTGCCT	TTGC	ACTGGG	TACG	CTGCC	AATC	TTTNGG
	551	CAAT	CGGCAT	TTTT	CCCTG	CAAC	TGNAAA	AAAT	CATGCA	AAAC	CGATAT
	601	ATCC	CGCTGT	GTAC	GGGATT	ATCC	GATATCA	TTAT	GGGCAT	TATG	GAACAT
	651	TGCG	GTCTGT	TGGC	TGTAA						

This encodes a protein having amino acid sequence <SEQ ID 396>:

45	1	MNX	DITFL	TLFLL	GFGGTH	CIGMCG	LSSAFALQLP	PHI	NRXW	LILLN	
	51	TCRV	SSYTAI	GLIL	GLIGQV	GVSLDQ	TRVX	QNIL	YTAAN	LLFL	GLYLS
	101	GISS	LAARKIE	KIKG	KPIWRNL	NPIL	NRLLP	IKSIP	ACLAVG	ILWGW	LFECGL
	151	VYS	ASLYALG	SGS	AATGGGLY	MLAF	ALGTLP	NLXA	IGIFSL	QLXKIM	QNR
	201	IRL	CTGLSVS	LWAL	WKLAVL	WL	*				

50 ORF103a and ORF103-1 show 97.7% identity in 222 aa overlap:

		10	20	30	40	50	60
	orf103a.pep	MNX	DITFL	TLFLL	GFGGTH	CIGMCG	LSSAFALQLP
	orf103-1	MNH	DITFL	TLFLL	GFGGTH	CIGMCG	LSSAFALQLP
		10	20	30	40	50	60
		70	80	90	100	110	120

orf103a.pep	GLILGLIGQVGVSLDQTRVXQNILYTAANLLLFLGLYLSGISSLAAKIEKIGKPIWRNL	
orf103-1	GLILGLIGQVGVSLDQTRVXQNILYTAANLLLFLGLYLSGISSLAAKIEKIGKPIWRNL	70 80 90 100 110 120
orf103a.pep	NPILNRLPIKSIIPACLAVGILWGWLPCGLVYSASLYALGSGSAATGGLYMLAFALGTLP	130 140 150 160 170 180
orf103-1	NPILNRLPIKSIIPACLAVGILWGWLPCGLVYSASLYALGSGSAATGGLYMLAFALGTLP	130 140 150 160 170 180
orf103a.pep	NLXAIGIFSLQLKKIMQNRVIRLCTGLSVSLWALWKLAVLWLX	190 200 210 220
orf103-1	NLLAIGIFSLQLKKIMQNRVIRLCTGLSVSLWALWKLAVLWLX	190 200 210 220

Homology with a predicted ORF from *N.gonorrhoeae*

ORF103 shows 95.5% identity over a 222aa overlap with a predicted ORF (ORF103.ng) from *N. gonorrhoeae*:

orf103.pep	MNHDITFLTLFLLGFFGGTHCIGMCGGLSSAFXXQLPPHINRFWLILLNTGRVSSYTAI	60
orf103ng	MNHDITFLTLFLLGFFGGTHCIGMCGGLSSAFALQLPPHINRFWLILLNTGRISSTYAI	60
orf103.pep	GLILGLIGQVGVSLDQTRVXQNILYTAANLLLFLGLYLSGISSLAAKIEKIGKPIWRNL	120
orf103ng	GLILGLIGQVGVSLDQTRVXQNILYTAANLLLFLGLYLSGISSLAAKIEKIGKPIWRNL	120
orf103.pep	NPILNRLPIKSIIPACLAVGILWGWLPCGLVYSASLYALGSGSAATGGLYMLAFALGTLP	180
orf103ng	NPILNRLPIKSIIPACLAVGILWGWLPCGLVYSASLYALGSGSAATGGLYMLAFALGTLP	180
orf103.pep	NLLAIGIFSLQLKKIMQNRVIRLCTGLSVSLWALWKLAVLWL	222
orf103ng	NLLAIGIFSLQLKKIMQNRVIRLCTGLSVSLWALWKLAVLWL	222

The complete length ORF103ng nucleotide sequence <SEQ ID 397> is:

1	ATGAACACG	ACATCACITT	CCTCACCGTG	TTCCTGCTCG	GTTTCTTCGG
51	CGGAACACG	TGCATCGGTA	TGTGCGGCGG	ATTAAAGCAGC	GC GTTTGCGC
101	TCCAACCTCC	CCGCATATC	AACCGCTTTT	GGCTGATTCT	GCTGCTTAAC
151	ACAGGACGGA	TAAGCAGCTA	TACGCGAATC	GGCCTGATGC	TCGATTATAT
201	CGGACCACTC	GGCATTTCAC	TCGACCAATC	CCGCTCTCTG	CAAAATATT
251	TATACAGAGC	CTCCACCTC	CTGCTGCTCT	TTCAGCGCT	ATACTTGAGC
301	GGTATTCTCT	CCTTGCGGCG	AAAATCGGAG	AAAATCGGCA	AACCGATATG
351	GGCRAACCTG	AACCGGATAC	TCAACCGCGT	GCTGCCCAT	AAATCCATAC
401	CCGCTGCTCT	TGCTGTCCGA	ATATTATGGG	GCTGCCCTGC	GTGCGGACTG
451	GTTCACAGCG	CATCACTTTA	CGCGCTGGGA	AGCGGTAGTG	CGACAACCGG
501	CGGACTGTAT	ATGCTTGCGT	TTGCACTGGG	TACGCTGCC	AATCTTTTGG
551	CAATCGGCAT	TTTTTCCCTG	CAACTGAAAA	AAATCATGCA	AAACCGATAT
601	ATCCCGCTGT	GTACAGGATT	ATCCGTATCA	TATGGGCAT	TATGGAAGCT
651	TGCCGCTCTG	TGGCTGTA			

This encodes a protein having amino acid sequence <SEQ ID 398>:

1	MNHDITFLTL	FLLGFFGGTH	CIGMCGGLSS	AFALQLPPI	NRFWLILLN
51	TGRISSTYAI	GLMLGLIGQL	GLSLDQTRVL	QNILYTAASL	LLLFLGLYLS
101	GISSLAARKIE	KIGKPIWRNL	NPILNRLPI	KSIIPACLAVG	ILWGWLECGSL
151	VYSASLYALG	SGSATTGGLY	MLAFALGTLP	NLLAIGIFSL	QLKKIMQNRV
201	IRLCTGLSVS	LWALWKLAVL	WL*		

In addition, ORF103ng and ORF103-1 show 97.3% identity in 222 aa overlap:

orf103-1.pep	MNHDITFLTLFLLGFFGGTHCIGMCGGLSSAFALQLPPHINRFWLILLNTGRVSSYTAI	10 20 30 40 50 60

orf103ng	MNHDTFTLTLFLLGGTTCIGMCGGLSSAFALQLPHINRFLILLNTGRISSTYAI	10	20	30	40	50	60
		70	80	90	100	110	120
5 orf103-1.pep	GLILGLIGVGVSVDQTRVLQNILYTAANLLLLFLGLYLSGSISSAAKIEKIGKPIWRNL						
orf103ng	GLMLGLIGLGLISLDQTRVLQNILYTAANLLLLFLGLYLSGSISSAAKIEKIGKPIWRNL	70	80	90	100	110	120
		130	140	150	160	170	180
10 orf103-1.pep	NPILNRLPIKSIPACLAVGILWGLVPCGLVYSASLYALSGSSAATGGLYMLAFALGTLF						
orf103ng	NPILNRLPIKSIPACLAVGILWGLVPCGLVYSASLYALSGSSAATGGLYMLAFALGTLF	130	140	150	160	170	180
		190	200	210	220		
15 orf103-1.pep	NLLAIGIFSLQLKKIMQNRVIRLCTGLSVSLWALWKLAVLWLX						
orf103ng	NLLAIGIFSLQLKKIMQNRVIRLCTGLSVSLWALWKLAVLWLX	190	200	210	220		

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 47

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 399>:

```

1  ATGGAAAAACC  AAAGGCGCGT  CCTAGGCTTT  CGCTTGGCAC  TTTTGGCGGC
51  GATGACGTTG  GGAACGCTGC  CGAT.TCCGT  GCGGCAGGTA  TTGAAGTTTG
101  TCGATGCGCC  GACGCTGGTG  TGGGTGCGTT  TTACCGTGGC  GCGCGCGGTA
151  TTGTTTGTTT  TGCTGGCACT  GGGCGGGCGG  CTGCCGAAGC  GGGCAGGATT
201  TTTCTTGGTG  CTCATTACGG  CTGCTGCTGC  TCGCGGTGGC  GGGCATTTCG
251  GCAAACTTTG  TGCTGATTGC  CCAAGGGCTG  CATTATATTT  CGCGGACCAC
301  GACGCAGGTT  TTTGGGCAGA  TTTGCGCGTT  TACGATGATT  GTWTCGGTGT
351  TGTGTGTGTT  TAAAGACCGG  ATGACTGCGC  CTCAGAAAT  CGGCTTGGTT
401  TTGCTGCTGT  CCGGTTTGGT  TAIGTATTTT  AACGATAAAT  TCGGCGAGTT
451  GTGCGGTTTG  GGGCGGTATG  C.AAGGGCGT  GTTGTGTGT  GGGCAGGCA
501  CTATGCGATG  GGTCTGTAA  GCGCTGCGCG  AAAGCTGCT  GTCCGGCCAA
551  TTGCGGCGCG  AACAGATTCT  GCTGTGTATT  TATCGCGCAA  GTGCCGCGCT
601  GTTCCGTGCG  TTTGCCGCA  CGGCACACAT  CGGAAGTATG  GACGGTACGT
651  TGGCGTGGGT  ATGTATTGG  TATTGCTGCT  TGAATACGTT  AATCGGTTAC
701  GGCTCGTTCG  GCGAGGCGTT  GAAACATTGG  GAGGCTTCCA  AAGTCAGCGC
751  GGTAAACAACC  TTGCTCCCG  TGTTTACCGT  AATAAATACT  TTGCTCGGGC
801  ATTATGTGAT  GCGTGAAACT  TTTGCCGCGC  CGGA..

```

This corresponds to the amino acid sequence <SEQ ID 400; ORF104>:

```

1  MENQRFLLGF  RLALLAAMTW  GLFXSVRQV  LKFVDAPTLV  WVRFTVAAAV
51  LFLVLLALGGR  LPKRRDFSWC  SFRLLLLGVA  GISANVFLIA  QGLHYISPTT
101  TQVLWQISPF  TMIVVGVLVF  KDRMTAAQKI  GLVLLLAGLL  MYFNDKFGEL
151  SELGAYXKGV  LLCAAGSMAW  VCNAAVQKLL  SAQFGPQQIL  LLIIYAASAAV
201  FLFPFAEPAHI  GMSDGTLAW  CIAYCCLNTL  IGYGSFGEAL  KHWEASKVSA
251  VTTLLEFVTV  INTLLGHYVM  PETFAAP...

```

Further work revealed further partial DNA sequence <SEQ ID 401>:

```

1  ATGGAAAAACC  AAAGGCGCGT  CCTAGGCTTC  CGCTTGGCAC  TTTTGGCGGC
51  GATGACGTTG  GGAACGCTGC  CGATTGCGGT  GCGGCAGGTA  TTGAAGTTTG
101  TCGATGCGCC  GACGCTGGTG  TGGGTGCGTT  TTACCGTGGC  GCGCGCGGTA
151  TTGTTTGTTT  TGCTGGCACT  GGGCGGGCGG  CTGCCGAAGC  GCGCGGATTT
201  TTCTTGTGTC  TCATTACGCG  TGCTGTGCT  CGGCGTGGCG  GGCATTTCGG

```

251 CAAACTTTGT GCTGATTGCC CAAGGCTGCG ATTATATTTT GCCGACCAGG
 301 ACGCAGGTTT TGTGSCAGAT TTGCGCCGTT ACGTAGATTG TTGTCGGTGT
 351 GTTGGTGTTT AAGACCGGA TGACTGCCCG TCAGAAANTC GCGTTGGTGT
 401 TCGCTGTTGG CGGTTTGCTT ATGTTTTTTA ACAGATAAAT CGCGAGGTGT
 451 TCGGCTTTGG CGCGGTATGC GAAGGGCGTG TTGCTGTGCTG CGCAGCGCAG
 501 TATGGCATGG GTGCTGTATG CCGTGGCGCA AAGCTGCTGT TCGCGCAAT
 551 TCGGCGCGCA ACAGATTCTG CTGTTGATT ATGCGGCAAG TGCCGCCGTG
 601 TTCCGTCCGT TTGCGCAACC GGCACACATC GGAAGTTTGG ACGGTACGTT
 651 GCGGTGGGTT TGTTTTGCGT ATTGCTGCTT GAATACGTTA ATCGGTACG
 701 GCTCGTTCGG CGAGGCGGTG AACATTGGG AGGCTTCCAA AGTCAGCGCG
 751 GTAAACACCT TGCTCCCGGT GTTTACCGTA ATAwTwwCTT TGCTCGGCGA
 801 TTAATGATG CCGTGAACCT TTGCGCGGCC GGA...

This corresponds to the amino acid sequence <SEQ ID 402; ORF104-1>:

1 MENQRPLLG F ALALLAAMTW GTLPVAVRQV LKFVDAPTLV WVRFTVAAAV
 51 LFVLLALGGR LPKRRDFSWC SFRLLLLGVA GISANFVLIA QGLHYISPTT
 101 TOVLWQISPF TMIVVGVLFV KDRMTAAQKI GLVLLLAGLL MFFNDKFGEL
 151 SGLGAYAGV LCAAGSMAW VCAYAQAQKL SAQFGPQOIL LLIIYASAAV
 201 FLFPFAEPAH GSLDGLTAW CFAYCCLNTL IGYGSGEAL KHWEASKVSA
 251 VTTLLPVFTV IXXLLGHYVM PETFAAP...

Computer analysis of this amino acid sequence gave the following results:

Homology with hypothetical HI0878 protein of *H. influenzae* (accession number U32769)

ORF104 and HI0878 show 40% aa identity in 277aa overlap:

orf104	4	QRPLLGFRLLAAMTWGTL	LPXSVRQVLK	FVDAPTLVWXXXXXXXXXXXXXXXXXXXXP-	62
		Q+PLLG AL+ AM WG+LP	+++QVL	++A T+VW	P
HI0878	3	QQLPLGFTT	FALITAMAGSL	LPALAKQVLSVMNAQTIVWVFIIAAVSL	LALLAYKKQLPE 62
orf104	63	--KRRDFSWCSFRLL	LLGVAGISANFVLIAQGLHYISPTT	TOVLWQISPTT	MIVVGVLFV 120
		K R ++W	+L+GV Q++NFI +	L+YI P+ Q+	+S F N++ GVL+P
HI0878	63	LMKVRQYAW----	IMLIGVIGLTSN	FLFSSSINLYIEPSVAQIFTHLSSFGMLIC	GVLFV 118
orf104	121	KDRMTAAQKIXXXXXXX	MYFNDKFGELSGLGAYXKGVLLCAAGSMAWVCNAVAQKLL	180	
		K+++ OKI	++FND+F	+GL Y GV+L	G++ WV +AQKL+
HI0878	119	KEKLGHLHQIGL	FLLLIGLGLFFNDRF	DAFAGL	INQYSTGVILGVGGALIVWAYGMAQKLM 178
orf104	181	SAQFGPQOILLII	YASAAVFLPFAEPAHIGSM	DMGTLAWVCIAYCCLNTL	IGYSGFGEAL 240
		+F QOILL++Y A	F+P A++ + +	LA +C	YCCCLNTLIGYSG+ EAL
HI0878	179	LKRFNSQOILL	MMYLGCIAFMFMA	DFSQVQELT-PLALICFIY	CCLNTLIGYSGYAEAL 237
orf104	241	KHWEASKVSAVTT	LLPVFTVINILLGHYVMP	ETFAAP	277
		W+ SKVS V	TL+P+FT++ +	+ HY P	FAAP
HI0878	238	NRWDVSKVSV	VITVPLFTILF	SHIAHYFSPADFAAP	274

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF104 shows 95.3% identity over a 277aa overlap with an ORF (ORF104a) from strain A of *N.*

meningitidis:

orf104.pep	10	20	30	40	50	60
	MENQRPLLGFRLLAAMTWGTL					
orf104a	MENQRPLLGFRLLAAMTWGTL					
	10	20	30	40	50	60
orf104.pep	70	80	90	100	110	120
	LPKRRDFSWCSFRLLLVAGISANFVLIAQGLHYISPTT					
orf104a	LPKRRDFSWCSFRLLLVAGISANFVLIAQGLHYISPTT					
	130	140	150	160	170	180
orf104.pep	KDRMTAAQKIGLVLLLAGLLMYFNDKFGELSGLGAYXKGVLLCAAGSMAWVCNAVAQKLL					

		130	140	150	160	170	180
5	orf104a	<div> </div> <div>KDRMTAAQKIGLVLLLAGLIMFNDKFGELSGLGAYAKGVLLCAAGSMAWVCYAVAKLL</div> <div> </div>					
	orf104.pcp	<div> </div> <div>SAQFGQQIILLIIYAAASAVLPFAFPAHIGSMGDTLAWVCYIAYCCINTLIGYSGFEAL</div> <div> </div>					
10	orf104a	<div> </div> <div>SAQFGQQIILLIIYAAASAVLPFAELAHIGSLDGTLAWVCFAYCCINTLIGYSGFEAL</div> <div> </div>					
	orf104.pcp	<div> </div> <div>KHWEASKVSAVTTLPVETVINTLLGHYVMPETFAAP</div> <div> </div>					
15	orf104a	<div> </div> <div>KHWEASKVSAVTTLPVETVIFSLGHYVMPDTEFAAPDMNGLYAGALVVGGAATTAAG</div> <div> </div>					
		250	260	270	280	290	300

The complete length ORF104a nucleotide sequence <SEQ ID 403> is:

	1	ATGGA AAAACC	AAAGCGCCGCT	CGTAGCCGCT	CGGCTGGGACG	TTTTGGCGGC
	51	GATGACGCTGG	GAAGCGCTGC	CGATTCCGCT	GCGGTCAGGTA	TTGAA GTTGT
20	101	TGCGATCGCG	TGCGGCTGTGT	TGGGTCGCGT	TTACCGCTGAG	GGCGGCGGTA
	151	TTTGTGCTTT	TGCTGCGCATT	GGGCGGCGCG	TGCGCGGAAT	GGCGGGATTT
	201	TTCTTGGTGTG	TGCTT CAGGC	TGCTGTCTGT	CGGCTGCGGG	GGCATTTCGG
	251	CAAACTTTTGT	GATTGATTGCC	CAAGCGCTCG	ATTATATTTC	GCCGACGAGT
	301	ACCGACGATTT	TGCGGGACAT	TTCGCGCTTT	ACGATGATTG	TGCTCGGCTGT
25	351	GTGGGTGTTT	AAAGACCCGGA	TGACTCGCGC	TGAGAAATCT	GGTTTGGT TT
	401	TGCGTCTTGC	CGGTTTGCCT	ATGTTT TTTA	ACGATAAAAT	GGCGCAGTTG
	451	TGCGGTTTGG	CGGCTGATGC	GAAGGCGCTG	TGCTGTGTGT	GCGGACGACG
	501	TATGGCATGG	GCTGATTTCG	CCGTGGCGCA	AAAGCTCGTC	TGCGCGCGAT
	551	TGCGGCGCGA	ACAGATTCTG	CTGTGTTAGT	ATGCGGCGAG	TGCGCGCGAT
30	601	TTCCCTGCGT	TGTCGGCACT	GGCACAATCT	GAAAGTTTGG	ACGGTACGTT
	651	GGCGGGTGTT	TTGTTTGGCT	ATTGCTGCTT	GAATACGTTA	ATCGGTGTTA
	701	GCTCGCTTGG	CGAGGCGTGT	CGCTCGCGCT	AGTACCTCGC	AGTACGCGG
	751	GTAAAGCTGG	CGCTCGCGCT	TTTACCGCTG	ATATTATTTG	TGCTCGGCGA
	801	TATGTGATGT	CGTAGTACTT	TTGCGGCGCC	GGATATGAAC	GGTTTGGGTT
35	851	ATGCGCGCGC	AGCTGCTGTG	CTGGGGGGTG	CGGTACCGCG	GGCGGTGGGG
	901	GACGAGCTGT	TCAAAGCCGG	CTAG		

This encodes a protein having amino acid sequence <SEQ ID 404>:

40

```
1  MENQRRLGGF ALALLAAWTW GTLPFAVRQV LKFVDAPTLV WVRFTVAABV
51  LFVLLALGGR LPKWRDPFSC KPRLLLLGVA GTSANFVLIA QGLHYTSPTL
101  TQWLQIQISPF TMIVUGVLVF KDRLMTAAQKI GLVLLLAGLL MFFNDKIFGEL
151  SGLGAYAKGV LILCAAGTMAW VCAYVAQKLL SQAFGQPQLL LLTYAASAAV
201  FLTPFABLAGH GSLDGLTMAV FCFFACVLTLL IYGSFGSEAL KHWAEKVSIA
251  VPTLLPVFTLV IFSLLGHYVM PDTFAAPDMN GLGYAGALVY VGGAVTAAVG
301  DRLFKRR*
```

ORF104a and ORF104-1 show 98.2% identity in 277 aa overlap:

45	orf104a.pep	10	20	30	40	50	60
	orf104-1	MENQRPLLGFAALLAAMTWGTLPIAVRQVLKFVDAPTLVWVRETVAAAVLFLVLLAIGR					
50	orf104a.pep	70	80	90	100	110	120
	orf104-1	LPKWRDFSWCSFRLLLLGVAGISANFVLIQAQGLHYISPTTTQVLQWISPTTMIWGVGLVF					
55	orf104a.pep	130	140	150	160	170	180
	orf104-1	KDRMTAAQKIGLVLLLAGLIMFNDKFGELSGLGAYAKGVLLCAAGSMWVCYAVAQKLL					
60	orf104a.pep	190	200	210	220	230	240
	orf104-1	KDRMTAAQKIGLVLLLAGLIMFNDKFGELSGLGAYAKGVLLCAAGSMWVCYAVAQKLL					

5	orf104a.pep	SAQFGPQQIILLIYAASAAVFLPFAELAHIGSLDGTLAWCVFAYCCINTLIGYGSFGEAL	
	orf104-1	SAQFGPQQIILLIYAASAAVFLPFAEPAHIGSLDGTLAWCVFAYCCINTLIGYGSFGEAL	
		190 200 210 220 230 240	
10	orf104a.pep	250 260 270 280 290 300	
	orf104-1	KHWEASKVSAVTTLLPVFTVIFSLLGHYVMPDTFAAPDMNGLGYAGALVVVGAVTAAGV	
		250 260 270	

Homology with a predicted ORF from *N.gonorrohaeae*

ORF104 shows 93.9% identity over a 277aa overlap with a predicted ORF (ORF104.ng) from *N.*

gonorrohaeae:

15	orf104.pep	MENQRPLLGFRLALLAAMTWGTLFXSVRQVLKFVDAPTLVWVRFTVAAAVLFVLLALGGR	60
	orf104ng	MENQRPLLGFRLALLAAMTWGTLPIAVRQVLKFVDAPTLVWVRFTVAAAVLFVLLALGGR	60
20	orf104.pep	LPKRRDFSWCSFRLLLLGVAGISANFVLIQAQGLHYISPTTTQVLWQISPTTMIIVGVLVF	120
	orf104ng	LPKRRDFSWHCSFRLLLLGVGTGISANFVLIQAQGLHYISPTTTQVLWQISPTTMIIVGVLVF	120
25	orf104.pep	KDRMTAAQKIGLVLLLAGLIMYFNDKFGELSGLGAYXKGVLCCAAGSMAWVCNAVAQKLL	180
	orf104ng	KDRMTAAQKIGLVLLVGLIMFNDKFGELSGLGAYAKGVLLCCAAGSMAWVCYAVAQKLL	180
30	orf104.pep	SAQFGPQQIILLIYAASAAVFLPFAEPAHIGSDMGTLAWCVIAYCCINTLIGYGSFGEAL	240
	orf104ng	SAQFGPQQIILLIYAASAAVFLPFAEPAHIGSLDGTLAWCVFAYCCINTLIGYGSFGEAL	240
35	orf104.pep	KHWEASKVSAVTTLLPVFTVINTELLGHYVMPETFAAP	277
	orf104ng	KHWEASKVSAVTTLLPVFTVIFSLLGHYVMPDTFAAPDMNGLGYAGALVVVGAVTAAGV	300

The complete length ORF104ng nucleotide sequence <SEQ ID 405> is predicted to encode a protein having amino acid sequence <SEQ ID 406>:

1	MENQRPLLGF	ALALLAAMTW	GTLPPIAVRQV	LKFVDAPTLV	WVRFTVAAAV
51	LFVLLALGGR	LPKRRDFSWH	SFRLLLLGVT	GISANFVLI	AQGLHYISPTT
101	TOVLWQISPF	TMIVGVVLVF	KDRMTAAQKI	GLVLLLVGLL	MFFNDKFGEL
151	SLGLAYAKGV	LLCAAGSMAW	VCYAVAQKLL	SAQFGPQQIL	LLIYAASAAV
201	FLXAEPAHI	GSLDGTLAW	CFVYCCNLTL	IGYGSFGEAL	KHWEASKVSA
251	VTTLLPVETV	IFSLLGHYVM	PDTFAAPDMN	GLGYVGLAV	VVGAVTAAGV
301	DRPFKRR*				

Further work revealed the complete gonococcal nucleotide sequence <SEQ ID 407>:

45	1	ATGGAACACC	AAAGCCCGCT	CCTAGGCTTC	GCGTTGGCAG	TTTGTGGCGC
	51	GATGACGCTG	GGGACGCTGC	CGATTGCCGT	GGCGCAGGTA	TTGAAGTTTG
50	101	TCGATGCGCC	GACGCTGCTG	TGGGTGCGTT	TTACCGTGGC	GCGCGCGGTA
	151	TTGTTTGCTT	TGCTGCGATT	GGCGGCGCGG	CTGCCGAGAC	GCGCGGATTT
55	201	TTCTTGCGAT	TCATTCAAGC	TGCTGCTGCT	CGCGCTGACG	GGCATTTCGG
	251	CAAACTTTGT	GCTGATTGCC	CAAGGCGTGC	ATTATATTTC	GCGGACACAG
60	301	ACGCAGGTTT	TGTGGCAGAT	TTCGCCGCTT	ACGATGATTG	TTGTCGCGCT
	351	GTTGGTGT	TTAAAGACCGA	tgactgctgcg	GCAGAAATC	GTTTGGTTT
65	401	TGCTGCTgtT	CGGTTgtCTT	ATGTTTttta	ACGACAAAT	CGCGAGTGT
	451	TCGGGTTTGT	GCGCGTATGC	GAAGGCGGTG	TTGCTGTGTG	CGCGAGGAG
70	501	TATGGCCTGG	GTGTGTTATG	CCGTGGCGCA	AAAGCTGCTG	TCGCGCAAT
	551	TCGGGCGCGA	ACAGATTCTG	CTGTGTGATT	ATGCGGcaag	tgccgcgGTG
75	601	TTCCtgccgt	TTGcgcaaac	GGCACACATC	GGAAAGTTtg	aCGGTACgct
	651	GGCGTGGGTT	TGTTTTGTGT	ATTGCTGCTT	GAATACGTTA	ATCGGTTACG
80	701	GCTCGTTCGG	CGAGGCGGTG	AAACATTGGG	AGGCTTCCAA	AGTCAGCGCG
	751	GTAACAACCT	TGCTCCCGCT	GTTACCGCTA	ATATTTTCTT	TGCTCGGGCA
	801	TTATGTGATG	CCTGATACCT	TTGCGCGGCC	GGATATGAAC	GTTTGGGTT

851 ATGTCGGCGC ACTGGTCGTG GTCGGGGGTG CGGTTACGGC GCGCGTGGG
 901 GACAGGCCGT TCAACGCCG CTAG

This corresponds to the amino acid sequence <SEQ ID 408; ORF104ng-1>:

5 1 MENQRPLLGG ALALLAAMTW GTLPFAVRQV LKFVDAPTLV WVRFTVAAAV
 51 LFVLLALGGR LPKRRDFSWH SFRLLLLGVT GISANFVLIA QGLHYISPTT
 101 TVQLWQISPTT TMIVVGVVLF KDRMTAAQKI GLVLLVGLL MFFNDKFGEL
 151 SGLGAYAKGV LLCAAGSMAW VCYAVALKLL SAQFGPQOIL LLIIYASAAV
 201 PLFPAEPAHI GSLDGTLAWV CFVYCCNLTL IGYGSFGEAL KHWEASKVSA
 251 VTTLLPVETV IFSLLGHYVM PDTFAAPDMN GLGYVGALVV VGGAVTAAVG
 10 301 DRFFKRR*

ORF104ng-1 and ORF104-1 show 97.5% identity in 277 aa overlap:

		10	20	30	40	50	60
orfl04-1.pep		MENQRPLLGFALALLAAMTWGTLPIAVRQVLKFVDAPTLVWVRFTVAAAVLFVLLALGGR					
orfl04ng-1		MENQRPLLGFALALLAAMTWGTLPIAVRQVLKFVDAPTLVWVRFTVAAAVLFVLLALGGR					
		10	20	30	40	50	60
orfl04-1.pep		70	80	90	100	110	120
orfl04ng-1		70	80	90	100	110	120
		130	140	150	160	170	180
orfl04-1.pep		KDRMTAAQKIGLVLLAGLMMFFNDKFGELSGLGAYAGVLLCAAGSMAWVCYAVALKLL					
orfl04ng-1		KDRMTAAQKIGLVLLAGLMMFFNDKFGELSGLGAYAGVLLCAAGSMAWVCYAVALKLL					
		130	140	150	160	170	180
orfl04-1.pep		190	200	210	220	230	240
orfl04ng-1		190	200	210	220	230	240
		250	260	270			
orfl04-1.pep		KHWEASKVSAVTTLLPVETVIFSLGLGHYVMPDTFAAP					
orfl04ng-1		KHWEASKVSAVTTLLPVETVIFSLGLGHYVMPDTFAAPDMNGLGYVGALVVVGGAVTAAVG					
		250	260	270	280	290	300

In addition, ORF104ng-1 shows significant homology with a hypothetical *H. influenzae* protein:

gi|1573895 (U32769) hypothetical [Haemophilus influenzae] Length = 306
 Score = 237 bits (598), Expect = 8e-62
 Identities = 114/280 (40%), Positives = 168/280 (59%), Gaps = 8/280 (2%)

45 Query: 30 QRPXXXXXXXXXXMTWGTLPFAVRQVLKFVDAPTLVXXXXXXXXXXXXXXXXXX- 88
 Q+P M WG+LPIA++QVL ++A T+VW P
 Sbjct: 3 QQPLLGFTFALITAMAWGSLPIALKQVLVSMNAQTIVWYRFIIAAVSLALLAYKKQLPE 62

50 Query: 89 --KRRDFSWHSFRLLLLGVTGISANFVLIAQGLHYISPTTQVLWQISPTTMIVVGVVLF 146
 K R ++Y ++LGV G++NF+L + L+YI E+ Q+ +S F M++ GVL+P
 Sbjct: 63 LMKVRQYAW---IMLIGVIGLTSNLLFSSSLNIEPSVAQIFIHLSSFGMLICGVLI 118

55 Query: 147 KDRMTAAQKIXXXXXXXXXXMMFFNDKFGELSGLGAYAGVLLCAAGSMAWVCYAVALKLL 206
 K+++ QKI +FND+F +GL Y+ GV+L G++ WV Y +AQKL+
 Sbjct: 119 KEKGLGHQIGLFLLLIGLGLFNDRFDAGLNQYSTGIVLGVGGALLWVAYGMAQKLM 178

60 Query: 207 SAQFGPQOILLIIYASAAVFLPFAEPAHIGSLDGTLAWVCFVYCCNLTLIGYGSFGEAL 266
 +F QOILL++Y A F+P A+ + + L LA +CF+YCCNLTLIGYGS+ EAL
 Sbjct: 179 LRKFNQQILLMMYLGCAIAFMADFSQVQELT-PLALICFIYCCNLTLIGYGSYAEAL 237

Query: 267 KHWEASKVSAVTTLLPVETVIFSLGLGHYVMPDTFAAPDMN 306
 W+ SKVS V TL+P+FT++FS + HY P FAAP++N

Sbjct: 238 NRWDVSKVSVVITLVPLFTILFISHIAHYFSPADFAPELN 277

Based on this analysis, including the presence of a putative leader sequence and several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 48

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 409>:

```

1  ATGGTAGCTC  GTCGGGCTCA  TAACCCGAAG  GTCGTAGGTT  CGAATCCTGT
51  CCCGCAACC  TAATTTCAAA  CCCCTCGGTT  CAATGCCGAG  GG.GTTTTGT
101 T.TTGCCTGT  TTCCGTGTTT  CTGTTTCTGT  CCGCTCCGTT  TTTTTCGGCG
151 ATTTTCTCTC  CGGCGCGCAAT  ATCGGAACGG  CAGACCCGCG  TCTGTTTGGC
201 GTTGCAAAAT  CAGGCAGTTT  GGCTACAATC  TTCCGCATTG  TCTTCAAGAA
251 AGCCAAACCAT  GCCGACCGCT  CGTTTTACCG  AATCCGTCAG  CAAACAAGAC
301 CTTGATGCTC  TGTTCGAGTG  GGCAAAAGCA  AGTTACGGTG  CAGAAAGTTG
351 CTGGAAGAAC  CTGATATCTGA  ACGGTCysCC  TTTGGGCAAC  CTGTCGCCGG
401 AATGGGTGGA  ACGCGTsmm  AAAGACTGGG  AGGCAGGCTG  CyCGGAGTCT
451 TCAGACGGCA  TTTTTCGTAA  TgCGGACGGC  TgGcCTGATA  TGGGCGGAcg
501 cTTACAGCAC  CTCGCCCTCG  GTTGGCACTG  TGCGGGGCTG  TTGGACGgST
551 GGCAGCAACG  GTGTTTCGAC  CTGACCGACG  CGCGCGCGCA  CCCCTGTGTC
601 ACGCTCGaAc  CGCGCSyTIT  mGCTCCTKTC  GGCATGCTCA  GCCCGCGCTC
651 CCATCTCAAC  GGTCTGCGCG  AATCGGACGG  CGATGGCAT  TCTGSGATAG
701 CGAGCGCGAG  TCCGCAACAA  GCAGTCGATC  CCAACAACAT  CGACAATACT
751 TCCGCGCGCG  CTGTTTCCGG  CGCGCAATG  CCGTCTGAAG  CCGGTGTGTCG
801 CGAAGCAGC  GAAGAACCG  GTTTGGATAA  AACGCTGcTT  CCGCTCATCC
851 GCCCGGTATC  GCAGCTGCAC  AGCCTCGCT  CCGTCAGCCG  GGGGTGACAC
901 AATGAAATCC  TGTATGTATT  CGATGCCGTC  CTGCCG...

```

This corresponds to the amino acid sequence <SEQ ID 410; ORF105>:

```

1  MVARRAHNPK  VVGSNFXPAT  XFQTPRENAE  XVXLXPFVSCF  LFPAASVFCCR
51  IFLPAALISER  QTAIVCLRLQI  QAVWLQSSAL  SSRKPTMPTV  RPTESVSKQD
101 LDALFEWAKA  SYGAESCWKT  LYLNKGXPLGN  LSPFWVERVX  KDWEAGCXES
151 SDGIFLNADG  WPDMGGRLLQ  LALGWHCAGL  LDGWRNECFD  LTDGGNNPLF
201 TLERAXRXPX  GLLSRAVHLN  GLTESDGRWH  FWIGRRSPHK  AVDPNKLDNT
251 XAGGVSGGEM  PSEAVCRESS  EEAGLDKTL  PLIRPVSQ  SLRSVSRGVH
301 NEILYVFDAV  LP...

```

Further work revealed the complete nucleotide sequence <SEQ ID 411>:

```

1  ATGCCGACCG  TCCGTTTTAC  CGAATCCGTC  AGCAACAAG  ACCTTGATGC
51  TCTGTTTCGAG  TGGGCAAAAG  CAAGTTACGG  TGCAGAAAGT  TGCTGAAAAA
101 CGCTGATPCT  GAACGCTGTG  CATTGCGCA  ACCTTCCGCC  GGAATGGGTG
151 CAGCGCTCA  AAAAAGACTG  GGAGCGAGCG  TCGTCGGAGT  CTTCAAGACG
201 CATTTTTCTG  AATCGGACG  GCTGCCCTGA  TATGGCGCGA  CGCTTACAGC
251 ACCTCGCCCT  CGGTGGCCAC  TGTGCGGGGC  TGTGAGACGG  CTGGCGCAAC
301 GAGTGTTTCG  ACCTGACCGA  CGGCGGGGCG  AACCCCTGT  TCACGCTOGA
351 ACGCGCGCGT  TTCCGTCCTT  TCGGACTGCT  CAGCCGCGCC  GTCCATCTCA
401 ACGGCTCGAC  CGAATCGGAC  GGCCGATGGC  ATTCTGGAT  AGGCAGGCGC
451 AGTCCGCACA  AAGCAGTCGA  TCCCAACAAA  CTCGACAMTA  CTGCGCGCGG
501 CGGTGTTTCC  GCGCGCGAAA  TGCCGCTGA  AGCCGTGTGT  CGCGAAGACA
551 GCGAAGAAGC  CGGTTTGGAT  AAAACGCTGC  TTCCGCTCAT  CGCCCGSGTA
601 TCGCAGCTGC  ACAGCGCTGC  CTCGCTCAGC  CGGGGTGTAC  ACAATGAANT
651 CCTGTATGTA  TTCGATGCCG  TCCTGCCCGA  AACCTTCTGT  CCTGAAATCT
701 AGGATGGCGA  AGTGGCGGGT  TTTGAGAAAA  TGGACATCGG  CGGTCTGTGT
751 GATGCCATGT  TGTGCGGAAA  CATGATGCAC  GACGCGCAAC  TGGTATGCTG
801 GGACGCGTIT  TGCCTGTAC  GTCTGATTGA  TGCCGCCCAT  CGGCTGTCCG
851 AGTGCGCTGA  CGGCATACGT  TTATAG

```

This corresponds to the amino acid sequence <SEQ ID 412; ORF105-1>:

1 MPTVRFTEVS SKQDLDALE WAKASYGAES CWKTLYLNGL PLGNLSPEWV
 51 ERVKKDWEAG CSESSDGIFL NADGWDFMG RQHLALGWH CAGLLDGNRN
 101 ECFDITDGGG NPLFTLERAA FRFFGLLSRA VHLNGLTESD GRWHFWIGRR
 151 SPHKAVDPNK LNTTACGVS GCEMPSEAVC RESSEEAGLD KTLPLIRPV
 201 SQLHSLRVS RGVHNEILYV FDAVLPEFVL PENQDGEVAG FEKMDIGGLL
 251 DAMLSGNMMH DAQLVTLDAF CRYGLIDAAH PLSEWLDGIR L*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF105 shows 89.4% identity over a 226aa overlap with an ORF (ORF105a) from strain A of *N.*

meningitidis:

	60	70	80	90	100	110
orf105.pep	ISERQTAVCLRLQIQAVWLQSSALSSRKPTMPTVRFTEVSWSKQDLDALEWAKASYGAES					
orf105a	MTPTVRFTEVSWKHDLDALEWAKASYGAES					
				10	20	30
orf105.pep	120	130	140	150	160	170
orf105a	CWKTLYLNGXPLGNLSPEWVERVKKDWEAGCXESSDGIFLNADGWDFMGGRRLQHLALGWH					
	40	50	60	70	80	90
orf105.pep	180	190	200	210	220	230
orf105a	CAGLLDGNWRNECFDLTDGGGNPLFTLERAXKRPXGLLSRAVHLNGLTESDGRWHFWIGRR					
	100	110	120	130	140	150
orf105.pep	240	250	260	270	280	290
orf105a	SPHKAVDPNKLDNTXAGGVSSEMPSEAVCRESSSEEAAGLDKTLPLIRPVSQLHSLRVS					
	160	170	180	190	200	210
orf105.pep	300	310				
orf105a	RGVHNEILYVDFDAVL					
	220	230	240	250	260	270
orf105a	RGVHNEILYVDFDAVLPEFVLPEFQDGEVAGFEKMDIGGLLAAMLSGNMMHDAQLVTLDAF					

The complete length ORF105a nucleotide sequence <SEQ ID 413> is:

1	ATGCCGACCG	TCCGTTTAC	CGAATCCGTC	AGCAAAACAG	ACCTTGATGC
51	CCTATTTCGAG	TGGGCAAAAGG	CAAGTTACGG	TGCGGAAAGT	TGCTGGAAAA
101	CGCTGTATCT	GAACGGTCTG	CCTTTGGGCA	ATCTGTCCGC	SGAATGGGCG
151	GAGGCGGTCA	AAAAAGACTG	GGAGGCGGCG	TGCTCGGAGT	CTTCAGACGG
201	CATTTCCTCG	AATGCGGACG	GCTGCGCCAG	TATGCGCAGA	CGCTTGCACG
251	ACCTCGCCCG	AATATCGGAA	GAACGGGACG	TGCTTCACGG	TGCTCGGACG
301	GAGTGTTTTCG	ACCTGACCGA	CGCGCGGACG	AATCCCTTGT	TGCGCTCGCA
351	ACGCGCGCGT	TTCCGTCCGT	TGCGACTGCT	CAGCGCGGCC	GTCCATCTCA
401	ACGGTTTGGT	CGAATCGGAC	GGCCGATGGC	ATTTCTGGAT	AGGCAGCGCG
451	AGTCCGCACA	AAGCAGTCGA	TCCCGACAAA	CTCGACATA	CTCGCGCGCG
501	CGGTGTTTTC	AGCGGTGAAT	TGCCGTCTGA	AACCGTGTGT	CGCGAAAGCA
551	GCGAAGAAGC	CGGTTTGGAT	AAAACGCTGC	TTCCGCTCAT	CGCGCGCGTA
601	TGCGAGCTGC	ACAGCCTGCG	CCCCGTGAGC	CGGGGTGTGC	ACAATGAAAT
651	CCTGTATGTA	TTCCGATGCG	TCTCGCCCGA	AACCTTCTGT	CCTGAAATAT
701	AGGATGCGCA	AGTGGCGGGT	TTTGAGAAAA	TGACATCGG	CGGTCTGTGT
751	GCTGCCATGT	TGTCGGGAAA	CATGATGCAC	GACCGGCAAC	TGGTTACGCT
801	GGACGCGTTT	TGCCGTTTAC	GTCTGATTGA	TGCGCGCCAT	CGGTGTCTCG
851	AGTGGCTGGA	CGGCATACGT	TTATAG		

This encodes a protein having amino acid sequence <SEQ ID 414>:

1 MPTVRFTEVS SKHDLDALE WAKASYGAES CWKTLVNLGL PLGNLSPEWA
 51 ERVKKDEWAG CSESSDGIPL NADGWPMGR RLQHLARIWK EAGLLHGWDR
 101 ECFDLTDGGS NPLFALERAA FRPFGLLSRA VHNLGLVESD GRWHFWIGRR
 151 SPHKAVDPDK LDNTAAGVSV SGELPSETVC RESSEAGLSD KTLPLIRPVP
 201 SQLHSRLPVS RGVHNEILYV FDAVLPETFL PENQDGEVAG FEKMDIGGLL
 251 AAMLSGNMMH DAQLVTLDAF CRYGLIDAAH PLEWLDGIR L*

ORF105a and ORF105-1 show 93.8% identity in 291 aa overlap:

		10	20	30	40	50	60
10	orf105a.pep	MPTVRFTEVS	SKHDLDALE	WAKASYGAES	CWKTLVNLGL	PLGNLSPEWA	ERVKKDEWAG
	orf105-1	MPTVRFTEVS	SKQDLDALE	WAKASYGAES	CWKTLVNLGL	PLGNLSPEW	VERVKKDEWAG
		10	20	30	40	50	60
		70	80	90	100	110	120
15	orf105a.pep	CSESSDGI	FLNADGWPM	GRRLQHLARI	WKEAGLLHGW	RCDFDLTDG	GSNPLFALERAA
	orf105-1	CSESSDGI	FLNADGWPM	GGRLQHLALG	WHCAGLLDGR	WNCEFDLTD	GGNPLFLTERAA
		70	80	90	100	110	120
20		130	140	150	160	170	180
	orf105a.pep	FRPFGLLS	RAVHNLGL	SVSDGRWHF	WIGRRSPHK	AVDPDKLDNT	AAGVSVSGELPSETVC
	orf105-1	FRPFGLLS	RAVHNLGL	TESDGRWHF	WIGRRSPHK	AVDPDKLDNT	AAGVSVSGGEMPEAVC
		130	140	150	160	170	180
25		190	200	210	220	230	240
	orf105a.pep	RESSEAGL	DKTLLPLIR	PVSQLHSRL	PVSRGVHNE	ILYVFDVLP	ETFLPENQDGEVAG
	orf105-1	RESSEAGL	DKTLLPLIR	PVSQLHSRL	SVSRGVHNE	ILYVFDVLP	ETFLPENQDGEVAG
		190	200	210	220	230	240
30		250	260	270	280	290	
	orf105a.pep	FEKMDIGGL	LAAAMLSGN	MMHDAQLVT	LDAFCRYGLI	DAAHPLSEW	LDGIRLX
	orf105-1	FEKMDIGGL	LDAAMLSGN	MMHDAQLVT	LDAFCRYGLI	DAAHPLSEW	LDGIRLX
		250	260	270	280	290	

Homology with a predicted ORF from *N.gonorrhoeae*

ORF105 shows 87.5% identity over a 312aa overlap with a predicted ORF (ORF105.ng) from *N.*

gonorrhoeae:

40	orf105.pep	MVARAHNPKV	VGSNXPXAT	XFTQFRFNA	EXVLKLPV	SCFLPFAAS	VFCRIFL	PAAISER	60
	orf105ng	MVARAHNPKV	VGSNFAPAT	KYQTFRFA	AGVLF----	FLPFAAS	VFCRIFL	PAAISER	55
45	orf105.pep	QTAVCLRLQ	IQAVWLQSS	ALSSRKPT	MPTVRFTE	SVSKQDLDA	LEWAKASY	GAESCWKT	120
	orf105ng	QAACVCLRL	QIQAVWLQ	SSALCSRK	PAMPTVRF	TESVSKQD	LDAFERAK	ASYGAES	115
50	orf105.pep	LYLNGXPL	GNLSPEW	VERVXKDW	EAGCXESS	DGIFLNAD	GWPDMG	GRRLQHLAL	180
	orf105ng	LYLNRLPL	GNLSPEWA	ERIKKDW	EAGCSESS	NGIFLNAD	GWPDMG	GRRLQHLAR	175
	orf105.pep	LDGWRNE	CFDLTDG	GGNPLFL	TLERAXXR	PXGLLSRA	VHNLGLT	ESDGRWHF	240
	orf105ng	LHGWRNE	CFDLTDG	GGNPLFL	TLERAAFR	PFGLLIRAV	HNLGLV	ESNGRWHF	235
55	orf105.pep	AVDPNKL	DNTXAGV	SGGEMPE	AVCRESE	EAGLDKTL	LPLIRPV	SRLSLR	300
	orf105ng	AVDPGKL	DNIAGGV	SGGEMPE	AVCRESE	EAGLDKTL	LPLIRPV	SRLSLR	295
60	orf105.pep	NEILYVFD	AVLP						312
	orf105ng	NEILYVFD	AVLPETFL	PENQDGE	VAGFEKMD	IGGLDAML	SKNMMH	DAQLVT	355

A complete length ORF105ng nucleotide sequence <SEQ ID 415> was predicted to encode a protein having amino acid sequence <SEQ ID 416>:

```

1  MVARRAHNPK  VVGSNFPAPAT  KYQTPRFNAE  GVLFLLFPAA  SVFCRIFLPA
5  51  AISERQAAVC  LRLQIQAVWL  QSSALCSRKP  AMETVRFTE  VSKQDLDALE
101  ERAKASYGAE  SCWKTLYLNR  LPLGNLSPEW  AERIKKDWEA  GCSESSNGIF
151  LNADGWDMG  RLQHLARTWN  NKAGLLHGWR  NECFDLTGG  GNPLFLTERA
201  AFRPGLLSRA  VHLNGLVESN  GRWHFWIGLR  RSPHKAVDPG  KLDNIAGGGV
251  SGGEMPSEAV  GEMPESEAVC  RESSEAGLD  KTLFPLIRPV  SRGVHNEILY
301  VFDAVLPETF  LPENQDGEVA  GFEMDIGGL  LDAMLSKNMM  HDAQLVTLDA
351  FYRYGLIDAA  HPLSEWLDGI  RL*

```

Further work revealed the complete nucleotide sequence <SEQ ID 417>:

```

1  ATGCCGACCG  TCGTGTTCAC  CGAATCCGTC  AGCAAAACAG  ACCTTGATGC
51  CCTGTTCCAG  CGGSCAAAAG  CAACTTACGG  TCCTGGAAGT  TCCTGGAARA
101  CGCTGTATCT  GACGCTCTTT  CTTTGGGCGA  ATCTGTCCGC  GGAATGGGCT
151  GAGCGCATCA  AAAAGACTGT  GGAGGCAGCG  TGCTCCGAGT  CTTCAGACGG
201  CATTTCCTCT  AATGCGGACG  GCTGCGCGGA  TATGGCGCGA  CGCTTCGACG
251  ACCTCGCCCG  CACATGGAAC  AAGCGGGGCG  TCGTTCACGG  ATGGCGCAAC
301  GAGTGTTCAG  ACCTGACCGA  CGGCGGGGCG  AACCCCTTGT  TCACGCTCGA
351  ACGCGCGCGT  TTCCGTCGCT  TCGGACTACT  CAGCCGCGCC  GTCCATCTCA
401  ACGGTTTGGT  CGAATCGAAC  GGCAGATGGC  ATTTTGGTAT  AGGCAGGCGC
451  AGTCGCGACA  AAGCAGTCGA  TCCCGGCAAG  CTCGACAATA  TTGCGGGCGG
501  CGGTGTTTCC  GCGCGCGAAA  TGCCGTCTGA  AGCGGTGTGC  CGCGAAAGCA
551  GCGAAGAAGC  CGGTTTGGAT  AAAACGCTGT  TTCGCTCAT  CGCGCCAGTA
601  TCGCGGCTGC  ACAGCCTTCG  CCCGCTCAGC  CGAGGTGTGC  ACAATGAATC
651  CCGTGTATGT  TTCGATGCCG  TCCTGCCCCA  AACCTTCTCT  CCGTGAATTC
701  AGGATGGCGA  GGTAGCGGGT  TTTGAAAAGA  TGGACATTGG  CGGCCTATTG
751  GATGCCATGT  TGTTCGAAAA  CATGATGCAC  GACGCGCAAC  TGCTTACGCT
801  GGACGCGTTT  TACCGTTACG  GTCTGATTGA  TCGCGCCCAT  CGCGTGTCCG
851  AGTGCGCTGA  CGGCATACGT  TTATAG

```

This corresponds to the amino acid sequence <SEQ ID 418; ORF105ng-1>:

```

1  MPTVRFTESV  SKQDLDALE  RAKASYGAES  CWKTLYLNR  LPLGNLSPEW
51  ERIKKDWEAG  CSESSDGIFL  NADGWDMG  RLQHLARTWN  KAGLLHGWRN
101  ECFDLTDGGG  NPLFLTERAA  FRFPGLLSRA  VHLNGLVESN  GRWHFWIGLR
151  SPHKAVDPGK  LDNIAGGGVS  GEMPESEAVC  RESSEAGLD  KTLFPLIRPV
201  SRLHSLRPVS  RGVHNEILY  VFDAVLPETF  LPENQDGEVA  FEKMDIGLL
251  DAMLSKNMMH  DAQLVTLDAF  YRYGLIDAAH  PLEWLDGIR  L*

```

ORF105ng-1 and ORF105-1 show 93.5% identity in 291 aa overlap:

		10	20	30	40	50	60
40	orf105-1.pep	MPTVRFTESVSKQDLDALEFAWKASYGAESCWKTLYLNLPLGNLSPEWVERVKKDWEAG					
	orf105ng-1	MPTVRFTESVSKQDLDALEFAWKASYGAESCWKTLYLNLPLGNLSPEWAERIKKDWEAG					
		10	20	30	40	50	60
		70	80	90	100	110	120
45	orf105-1.pep	CSESSDGIFLNADGWDMGGRQLQHLALGWHCAGLLDGNWNECFDLTDGGGNPLFLTERAA					
	orf105ng-1	CSESSDGIFLNADGWDMGGRQLQHLARTWNKAGLLHGWRNECFDLTDGGGNPLFLTERAA					
		70	80	90	100	110	120
50	orf105-1.pep	FRFPGLLSRAVHLNGLTESDGRWHFWIGRRSPHKAVDPNKLNDTAAGGVSGGEMPSEAVC					
	orf105ng-1	FRFPGLLSRAVHLNGLVESNGRWHFWIGRRSPHKAVDPKLDNIAGGGVSGGEMPSEAVC					
		130	140	150	160	170	180
55	orf105-1.pep	RESSEAGLDKTLFLIRPVSQLHSLRPVSRGVHNEILYVFDAVLPETFLPENQDGEVAG					
	orf105ng-1	RESSEAGLDKTLFLIRPVSRHSLRPVSRGVHNEILYVFDAVLPETFLPENQDGEVAG					
		190	200	210	220	230	240
60		190	200	210	220	230	240

```

                250      260      270      280      290
orf105-1.pep    FEKMDIGGLDAMLDSGNMMHDAQLVTLDAFCRYGLIDAAHPLESWLDGIRLX
orf105ng-1      FEKMDIGGLDAMLDSKNMMHDAQLVTLDAFTRYGLIDAAHPLESWLDGIRLX
                250      260      270      280      290

```

Furthermore, ORF105ng-1 shows homology with a yeast enzyme:

```

sp|P41888|TNR3_SCHP1 THIAMIN PYROPHOSPHOKINASE (TPK) (THIAMIN KINASE)
>gi|1076928|pir|t852350 thiamin pyrophosphokinase (EC 2.7.6.2) - fission yeast
[Schizosaccharomyces pombe] >gi|666111(X84417) thiamin pyrophosphokinase
[Schizosaccharomyces pombe] >gi|2330852|gnl|PID|e334056 (Z98533) thiamin
pyrophosphokinase [Schizosaccharomyces pombe] Length = 569
Score = 105 bits (259), Expect = 4e-22
Identities = 64/192 (33%), Positives = 94/192 (48%), Gaps = 3/192 (1%)

```

Query: 268 NKAGLLHGWRNECFDLTDGGGNPLFTLERAARFRPGLLSRAVHLNGLVESNGRW--HFWI 441
N G+ WRNE ++ P+ +ER F FG LS VH ++ W+

Sbjct: 96 NTFGIADQWRNELYTVYGSKKKPVLAVERGGFWLFGLSTGVHCTMYIPATKEHPLRIW 155

Query: 442 GRRSPHKAVDPGKLDNIAGGGVSGGEMPSEAVCRESSSEAGLDKTLFPLIRPVSRHLSLR 621
RRSP K P LDN GG++ G+ + +E SSEA LD + LI P + ++
Sbjct: 156 PRRSPTKQTWPNYLDNSVAGGIAHGDSVIGTMIKEFSEANLDVSSMNLII-PCGTVSYIK 214

Query: 622 PVSRG-VHNEILYVFDAVLPEFLPENQDGEVAGFEKMDIGLLDAMLSKNMMHDAQLVT 798
R + E+ YVFD ++ +P DGEVAGF ++ +L + K+ + LV
Sbjct: 215 MEKRHWIQPELQYVFDFLPVDDLVIPIRINDGEVAGFSLLPLNOVLHELELKSFKPNCALVL 274

Query: 799 LDAFYRYGLIDAAHP 843
LD R+G+I HP
Sbjct: 275 LDELIRHGIITPOHP 289

Based on this analysis, including the presence of a putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 49

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 419>:

1	ATGAATAGAC	CCAGACCAACC	CTCTTCCTCG	CCCGAAGTCG	CGGTTCGCGG
5	CAACCAACGC	CTGACGGGTGA	AAGTGAATCT	CACACGACGC	TGTGTCAATT
101	TCGATATGGAC	GACATTTCGA	TGATATTCTG	CCGTATTGAT	TATCTCTGTT
1	CCGATTATTG	TGACTACTAC	CAGGAARACG	ACCTTGGAGG	GACAAATTTT
203	CGGATGATG	CGGATGATG	CGGATGATG	CGGATGATG	GACACATATG
251	CAGCGAATAT	CTCTGAGAAT	GGGAAATGCG	CGGATGATG	CGGATGATG
301	TTCGCGGTTT	CGACCTCAGC	TTTCCGCGCA	GAGGTTAGCG	TGCACGACGA
351	GTGGAARACG	GAGCGGACAT	TGAAGAAATC	TCGTCAGACA	CAGGACATCG
401	GTCTGCTGAA	CTGTATACAC	GGGAATGAAA	CGCGGACCT	TATGAAACAT
451	CGGACGCTT	TCGGAACACA	GGAACTCCAT	ATTTCGCAC	AAGATAGACG
501	TCGACGATG	TCGCGGAGCA	TCGCGGAGCA	AATGTTCGAC	GAATATCGTT
551	TCGATATGCG	CAATTA			

This corresponds to the amino acid sequence <SEO ID 420: ORF107>:

```

1  MNRPKQPFPR PEVAVARQTS LTGKVILTRP LSFSLWTTFA SISALLIILF
51  LIFGNYTRKT TVEGQQLPQS GVIRVYAPDT XTITAKFVED GXKVKAGDKL
101  FALSTSRRGA GGSVQQLQLT EAVLKKTLAE QYELGRLLKH GNETRSLKAT
151  VERTENQELH ISQOIGOKR RIRLAEMLO KRFRLSXO*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF107 shows 97.8% identity over a 186aa overlap with an ORF (ORF107a) from strain A of *N.*

meningitidis:

5	orf107.pep	MNRPKQFFRPEVAVARQTSLTGKVLTRPLSFLSWTTTFASISALLIILFLIFGNYTRKT	10	20	30	40	50	60
	orf107a	MNRPKQFFRPEVAVARQTSLTGKVLTRPLSFLSWTTTFASISALLIILFLIFGNYTRKT	10	20	30	40	50	60
10	orf107.pep	TVEGQILPASGVIRVYAPDTXTITAKFVEDGKVKAGDKLFALSTSRFGAGGSVQQQLKT	70	80	90	100	110	120
	orf107a	TVEGQILPASGVIRVYAPDTXTITAKFVEDGKVKAGDKLFALSTSRFGAGGSVQQQLKT	70	80	90	100	110	120
15	orf107.pep	EAVLKKTLAEQELGRKLHIGNETRSKATVERLENQELHISQQIDGQKRRLAEEMLQ	130	140	150	160	170	180
	orf107a	EAVLKKTLAEQELGRKLHIGNETRSKATVERLENQELHISQQIDGQKRRLAEEMLQ	130	140	150	160	170	180
20	orf107.pep	KYRFLSXQX	189					
	orf107a	KYRFLSANDAVFKQEMMNVAELLEQKAKLDAYRREEVGLLQEIQTQNLTLXSLPQAA	190	200	210	220	230	

The complete length ORF107a nucleotide sequence <SEQ ID 421> is:

30	1	ATGAATAGAC	CAAAGCAACC	NTTCTTCGGT	CCCGAAGTCG	CGGTGCGCCG
	51	CCAAACACAGC	CTGACGGGTA	AAGTGATCTC	GACACGACCG	TGTGCTATTT
35	101	CCCTATGAC	GACATATATG	CGTATATGAT	TATCCTGTTT	
	151	TTGATATTGG	GAAACTATAC	CGCAAGACGA	ACAATGGAGC	GACAATTTTT
40	201	ACCTGCATCG	GGCGTAATCA	GGTGATATGC	ACCGGATACG	GGGACAATTA
	251	CNCGCAAAAT	CNTGGAAGAT	GGAGAAAAGG	TTAAGCTGGG	CGACAAGCTA
45	301	TTTGGCGCTT	CGACCTCACG	TTTCGGCGCA	GGAGATAGCG	TGCAGCAGCA
	351	GGTGA AAAACG	GAGGCAGTTT	TGAAGAAAAC	GTTGGCAGAA	CAGGAACCTG
50	401	GTCGTCTGAA	GCTGATACAC	GGGAATGAAA	CGCGCAGCCT	TAAAGCAACT
	451	GTCGAACGTT	TGGA AAAACCA	GGAACCTCAT	ATTTGCGAAC	AGATAGACGG
55	501	TCAGAAAAGG	CGCATTAGAC	TTGCGGAAGA	AATGTTGCAG	AAATATCGTT
	551	TCCTATCCGC	CAATGATGCA	GTGCCAAAAC	AAGAATGATG	GAATGTCAAG
60	601	GCAGAGCTTT	TAGAGCAGAA	AGCCAAAATT	GATGCCTACC	GCCGAGAAGA
	651	AGTCGGGCTG	CTTCAGGAAA	TCCGCACGCA	GAATCTGACA	TTGGNNAGCC
	701	TCCCCCAAGC	GGCATGA			

This encodes a protein having amino acid sequence <SEQ ID 422>:

45	1	MNRPKQFFR	PEVAVARQTS	LTKVILTRP	LSFSLWTTFA	SISALLIILF
	51	LIENYTRKT	TVEGQILPAS	GVIRVYAPDT	GTITAKFVED	GEKVKAGDKL
50	101	FALSTSRFGA	GDSVQQQLKT	EAVLKKTLAE	QELGRKLH	IGNETRSKAT
	151	VERLENQELH	ISQQIDGQKR	RIRLAEMMLQ	KYRFLSANDA	VPKQEMMNVK
	201	AELLEQKAKL	DAYRREEVGL	LQEIQTQNL	LKSLPQAA*	

Homology with a predicted ORF from *N.gonorrhoeae*

ORF107 shows 95.7% identity over a 188aa overlap with a predicted ORF (ORF107.ng) from *N.*

gonorrhoeae:

50	orf107.pep	MNRPKQFFRPEVAVARQTSLTGKVLTRPLSFLSWTTTFASISALLIILFLIFGNYTRKT	60
	orf107ng	MNRPKQFFRPEVAVARQTSLTGKVLTRPLSFLSWTTTFASISALLIILFLIFGNYTRKT	60
55	orf107.pep	TVEGQILPASGVIRVYAPDTXTITAKFVEDGKVKAGDKLFALSTSRFGAGGSVQQQLKT	120

```

orf107ng      TMEGQILPASGVIRVYAPDTGTITAKFVEDGEKVKAGDKLFALSTSRFGAGGSVQQQLKT 120
orf107.pep    EAVLKKTLEAQELGRKLKHENETRSLKATVERLENQELHISQQIDGQKRRIRLAEEMLQ 180
5             |||||||
orf107ng      EAVLKKTLEAQELGRKLKHENETRSLKATVERLENQELHISQQIDGQKRRIRLAEEMLR 180
orf107.pep    KYRFLSXQ 188
              |||||
orf107ng      KYRFLSAQ 188

```

The complete length ORF107ng nucleotide sequence <SEQ ID 423> is predicted to encode a protein having amino acid sequence <SEQ ID 424>:

```

1  MNRPKQPFRR PEVAIARQTS LTGKVLTRP LSFSLWTTFA SISALLIILE
51 LIFGNYTRKT TMEGQILPAS GVIRVYAPDT GTITAKFVED GEKVKAGDKL
101 FALSTSRFGA GGSVQQQLKT EAVLKKTLEA QELGRKLKH ENETRSLKAT
151 VERLENQKLH ISQQIDGQKR RIRLAEEMLR KYRFLSAQ*

```

Based on the presence of a putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 50

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 425>:

```

1  ATGCTGAATA CTTTTTTTGC CGTATTGGGC GGCTGCCTGC TGCT.TTGGC
51 GTGCGGCAAA TCCGTAARATA CGCGGTATCA GCGCAAAAC GCGGTACAAA
101 GCGCGCCGAA ACCGGTTTTC AAGCTCATAT ATATCGACAA TACGGCGATT
151 GCGCGTTTGG ATTTGGGACA AAGCAGCGAA GCGCAAAACCA ACGACGCGCAA
201 AAAACAAATC AGTTATCCGA TTAAGGCTTT GCGGGAACAA AATGTTATCC
251 GACTGATCGG CAAGCATCCC GGCGACTTGG AAGCGTCA GCGCAATGT
301 ATGGAACCCG ATGATAAGGA CAGTCGGGCA GATTGGGCAG AAAACGGCGT
351 GTGCCATACC TTGTTTGCCA AACTGGTGGG CAATATCGCC GAAGACGGCG
401 GCAAACTGAC GGATTACCTA GTTTCGCATG CCGCCTGCA ACCCTATCAG
451 GCAGGCAAAA GCGGCTATGC CGCGTGCAG AACGACGCT ATGTGCTGGA
501 AATCGACAGC GAAGGGGCGT TTTATTCCG CCGCGGCAT TATTGA

```

This corresponds to the amino acid sequence <SEQ ID 426; ORF108>:

```

1  MLNTFFAVLG GCLLXLPCGK SVNTAVQPQN AVQAPKPVF KVIYIDNTAI
51 AGLDLGQSSE GKTNDGKKQI SYPIKGLPEQ NVIRLIGKRP GDLEAVSGRC
101 METDDKDISPA GWAENGVCHT LFAKLVGNIA EDGKGLTDYL VSHAALQPYQ
151 AGKSGYAAVQ NGRYVLEIDS EGAFYFRRRH Y*

```

Further work revealed the following DNA sequence <SEQ ID 427>:

```

1  ATGCTGAAAA CATCTTTTGC CGTATTGGGC GGCTGCCTGC TGCTTGCGCG
51 CTGCGCGCAA TCGGAAARATA CGCGGAACA GCGCAAAAC GCGGTACAAA
101 GCGCGCCGAA ACCGGTTTTC AAGCTCAAAT ATATCGACAA TACGGCGATT
151 GCGCGTTTGG ATTTGGGACA AAGCAGCGAA GCGCAAAACCA ACGACGCGCAA
201 AAAACAAATC AGTTATCCGA TTAAGGCTTT GCGGGAACAA AATGTTATCC
251 GACTGATCGG CAAGCATCCC GGCGACTTGG AAGCGTCA GCGCAATGT
301 ATGGAACCCG ATGATAAGGA CAGTCGGGCA GATTGGGCAG AAAACGGCGT
351 GTGCCATACC TTGTTTGCCA AACTGGTGGG CAATATCGCC GAAGACGGCG
401 GCAAACTGAC GGATTACCTA GTTTCGCATG CCGCCTGCA ACCCTATCAG
451 GCAGGCAAAA GCGGCTATGC CGCGTGCAG AACGACGCT ATGTGCTGGA
501 AATCGACAGC GAAGGGGCGT TTTATTCCG CCGCGGCAT TATTGA

```

This corresponds to the amino acid sequence <SEQ ID 428; ORF108-1>:

```

1  MLKTSFAVLG GCLLLAACGK  SENTAEQPN  AVQSAPKPVF  KVKYIDNTAI
51  AGLLDQGSS E GRTNDGKKVT  SYPIKGLPEQ NVIRLIGKHP  GDLEAVSGKC
101  METDKDSDPA  GWAENGVCIT  LFAKLVGNI A EDGGLTDL  VSHAALQPYQ
151  AGKSGYAALV  NGRYVLEIDS  EGAFYFRRRH Y*

```

5 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.gonorrhoeae*

ORF108 shows 88.4% identity over a 181aa overlap with a predicted ORF (ORF108.ng) from *N. gonorrhoeae*:

10	orf108.pep	MLNTFFAVILGGCLLXLPCGKSVNTAVQPQNAVQSAKPVFKVIYIDNTAIGLDLQGSSE	60
	orf108.ng	MLKIFPAVILGGCLLLAAGCKSENTAEQPQNAQSAKPVFKVKYIDNTAIGLALQGSSE	60
15	orf108.pep	GKTNDGKKQISYPKLGPEQNVRLIGKHGPDLEAVSGKCMETDDKDSPAGWAENGVCHT	120
	orf108.ng	GKTNDGKKQISYPKLGPEQNAVRLTGKHNDELAVVGKCMETDGDKDESPAGWAENGVCHT	120
	orf108.pep	LFKALVGNIAEDGGKTDYLIHSALQPVQAGKSGYAAVQNGRYVLEIDSEGAFYFRRHHY	181
	orf108.ng	LFKALVGNIAEDGGKTDYLIHSALQPVQAGKSGYAAVQNGRYVLEIDSEGAFYFRRHHY	181

ORF108-1 shows 92.3% identity with ORF108ng over the same 181 aa overlap:

	orf108-1.pep	MLKTSFVAVLGCGCLLAAAGCKSENTAEPQONQVASAPKPVFKVKYINDNTAIGLDIGQSSE	60
25	orf108ng-1	MLKIFFAVLGCGCLLAAAGCKSENTAEPQONAAQSAPKPVFKVKYINDNTAIGLAGLIGQSSE	60
	orf108-1.pep	GKTNDGKKQISYPKIGLPEQNVIRLIGKHGPDLEAVSGKCMETDDKDSPAGWAENGVCHT	120
	orf108ng-1	GKTNDGKKQISYPKIGLPEQNAVRLTGKHPDLEAVSGKCMETDGGKDAQSGWAENGVCHT	120
30	orf108-1.pep	LFKAVIGNTAIEDGGKITYLVLSHALOPFYQAGKSGYAAVQNGRYVLEIDSEGAFYFRRHHY	180
	orf108ng-1	LFKAVIGNTAIEDGGKITYLVLSHALOPFYQAGKSGYAAVQNGRYVLEIDSEGAFYFRRHHY	181

The complete length ORF108ng nucleotide sequence <SEQ ID 429> is:

35	1	ATGCTGAAAA	tacctTTTGG	CGGTGTTgggg	ggCTgcotGC	TGCTTGCGCGC
	51	CTGGCGCAAA	TCGCAAAATa	cgggcgaaCA	CGCGCAAAAT	cgggCACAAA
	101	GCGCGCGGAA	ACCGGTTTTC	AAGATCAAA	ACATTCGACA	TACGCGGATT
	151	GCGCGGTTTGG	CTTTGGGACA	AAGTAGCGAA	AGGAAACCA	acgacgCGAA
	201	AAAACAAAATC	AGTTATcggA	TTAAGSGCTT	CGCGGAACAA	Aacgcgcctc
40	251	gCTGACCGG	AAGCATCC	AAGCATCTTG	Aagcgctcgt	CGCAAAATGT
	301	ATGGAACAACG	ACCGAAGA	GCGCGCTTCG	GGTTCGCGG	GAAGACGCGCT
	351	TGCGCATACG	TTTGTTCGCA	ACATGTGGG	CATATCGCG	AAAAGCGCGC
	401	CGAATACGCT	CGGCTTTCG	ATTTCGAT	CGCTGCTG	ACGCTGCTG
	451	CGAGGCAAA	CGCGCTATCG	CCGCTGCGAG	ACGACGACCT	ATGTGCTGGA
	501	AATCGACAGC	GagggGGCGT	TTTATtccg	agcgccgcct	tatttga

45 This encodes a protein having amino acid sequence <SEQ ID 430>:

```

1  MLKIPFAVLG  GCLLLAACGK  SENTAEQPQN  AAQSAPKPVF  KVKYIDNTAI
51  AGLALGOSSE  GKTNDGKKQI  SYPIKGLPEQ  NAVRLTGKHP  NDLEAVVGKC
101  METGDGKAPS  GWAENGVCHT  LFAKLVILNA  EDGGLTDYLL  ISHSALQPYQ
151  AKGSGYAAOV  GNRGYVLEDS  EGAFVFRHRH  Y*

```

50 Based on this analysis, including the presence of a predicted prokaryotic membrane lipoprotein lipid attachment site (underlined) and a putative ATP/GTP-binding site motif A (P-loop, double-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and

N.gonorrhoeae, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 51

The following DNA sequence was identified in *N.meningitidis* <SEQ ID 431>:

```

5      1  ATGGAAGATT  TATATATAAT  ACTCGCTTTG  GGTTTGGTTG  CGATGATTGC
      51  CGGATTTATC  GATGcgattg  cGggCGGGG  TGGTTTGATT  ACGCTGCCCG
     101  CACTCTTTGT  GGCAGGTATT  CCTCCGCTG  CGGCAATTGC  CACCACAAAG
     151  CTGCAAGCAG  CCGCTGCTAC  GTTTTCAGCT  ACGGTTTCTT  TTGCACGCCAA
     201  AGGTTTGATT  GATTGGAAGA  AAGTCTCCCC  GATTGCCGCA  GCATCGTTTG
     251  TAGGCGGCGT  GGCGGTGCA  TTATCGGTCA  GCTTGGTTTC  CAAAGATATT
     301  CTGCTgCGG  TCGTGCCGGT  TTTGTTGATA  TTTGTCCGAC  TGTATTTTGT
     351  GTTTTCGCC  AAGCTCGACG  GCAGTAAGGA  AGGCAAGGCC  AGAATGTCTT
     401  TTTTCTGTT  cGGGCTGACG  GTCGC.ACCG  CTTTGGGTT  TTTACGACGG
     451  TGTGTTCGGA  CCGGCTGTCG  GTCGTTTTT  TCTGATTGCC  TTTATGTTT
     501  TGCTCGGCTG  CAAgCTGTTG  AACGCGATGT  CTACACCAA  ATTGGCGAAC
     551  GTTGCTGCA  ATCTGGTTC  GCTATCGGTA  TTCCTGCTCG  ACGGTTTCAT
     601  TATTTTCCG  ATTGCGGCAA  CGaTGGCGGT  CGGTGCGGTT  GTCGGTGCGA
     651  ATTTAgGTGC  GAGATTTGCC  GtaCgctTCG  GTTCGAAGCT  GATTAA
  
```

This corresponds to the amino acid sequence <SEQ ID 432; ORF109>:

```

20     1  MEDLYIILAL  GLVAMIAGFI  DAIAGGGGLI  TLPALLLAGI  PPVSAIATNK
      51  LQAAATFSA  TVSFARKGLI  DWKGLPIAA  ASFVGGVAGA  LSVSLVSKDI
     101  LLAVVPELLI  FVALYFVSP  KLDGSKEGKA  RMSFFLFLGT  VXTAFGLRLR
     151  CVRTGCRLVF  SDCLYCFARL  QAVRDLVHQ  IGERCLQSWF  AIGPIAARFD
     201  YFPDCGNDGG  RCVCRCEFR  EICRTLFEA  D*
  
```

Further work revealed the following DNA sequence <SEQ ID 433>:

```

30     1  ATGGAAGATT  TATATATAAT  ACTCGCTTTG  GGTTTGGTTG  CGATGATTGC
      51  CGGATTTATC  GATGCGATTG  CGGGCGGGGG  TGGTTTGATT  ACGCTGCCCG
     101  CACTCTTTGT  GGCAGGTATT  CCTCCGCTGT  CGGCAATTGC  CACCACAAAG
     151  CTGCAAGCAG  CCGCTGCTAC  GTTTTCAGCT  ACGGTTTCTT  TTGCACGCCAA
     201  AGGTTTGATT  GATTGGAAGA  AAGTCTCCCC  GATTGCCGCA  GCATCGTTTG
     251  TAGGCGGCGT  GGCGGTGCA  TTATCGGTCA  GCTTGGTTTC  CAAAGATATT
     301  CTGCTGCGG  TCGTGCCGGT  TTTGTTGATA  TTTGTCCGAC  TGTATTTTGT
     351  GTTTTCGCC  AAGCTCGACG  GCAGTAAGGA  AGGCAAGGCC  AGAATGTCTT
     401  TTTTCTGTT  cGGGCTGACG  GTCGCACCGC  TTTTGGTTT  TTACACGGGT
     451  GTGTTGCGAC  CGGCTGTCGG  CTCGTTTTT  CTGATGCGCT  TTTATGTTTT
     501  GCTCGGCTCG  AAGCTGTTGA  ACGCGATGTC  TTACACCAA  TTGGCGAAGC
     551  TTGCTGCGAA  TCTTGGTTTC  CTATCGGTAT  TCCGTGTCGA  CGGTTTCGAT
     601  ATTTCCCGSA  TTTGCGCAAC  GATGCGCGTC  GGTGCGTTTG  TCGGTGCGAA
     651  TTTAGTCCG  AGATTTCGG  TCGGCTTGG  TCGAAGCTG  ATTAAGCCCG
     701  TCGTATCTCT  CACTACGATT  TCGATGGCTG  TCAAAATGTT  GATGACCGAG
     751  AGAAATCCGC  TGTATCAGAT  GATTGTTTCG  ATGTTTAAA
  
```

This corresponds to the amino acid sequence <SEQ ID 434; ORF109>:

```

45     1  MEDLYIILAL  GLVAMIAGFI  DAIAGGGGLI  TLPALLLAGI  PPVSAIATNK
      51  LQAAATFSA  TVSFARKGLI  DWKGLPIAA  ASFVGGVAGA  LSVSLVSKDI
     101  LLAVVPELLI  FVALYFVSP  KLDGSKEGKA  RMSFFLFLGT  VAPLLFYDGG
     151  VFGPGVGSFF  LIAFIVLLGC  KLLNAMSYKL  LANVACNLGS  LSVFLHLSGI
     201  IFPIAATMAV  GARFGANLGA  RFAVRFGSKL  IKPLLVISVI  SMAVKLLIDE
     251  RNPLQMIVS  MF*
  
```

Computer analysis of this amino acid sequence gave the following results:

50 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF109 shows 95.9% identity over a 147aa overlap with an ORF (ORF109a) from strain A of *N.meningitidis*:

		10	20	30	40	50	60
5	orf109.pep	MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPFVSAIATNKLQAAATFSA					
	orf109a	MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPFVSAIATNKLQAAATFSA					
		10	20	30	40	50	60
	orf109.pep	TVSFARKGLIDWKKGLPIAAASFVGGVAGALSVSLVSKDILLAVVPVLLIFVALYFVFSP					
10	orf109a	TVSFARKGLIDWKKGLPIAAASFVGGVAGALSVSLVSKDILLAVVPVLLIFVALYFVFSP					
		70	80	90	100	110	120
	orf109.pep	KLDGSKGKARMSFFLFGTLVXTAFGLRRRCVTRGCRVFSDCCLYCFARLQAVRDVLHQ					
15	orf109a	KLDGSKGKARMSFFLFGTLVAPLLGFDYGVFGVGSFFLIAFIVLLGCKLLNAMS YTK					
		130	140	150	160	170	180

The complete length ORF109a nucleotide sequence <SEQ ID 435> is:

20	1	ATCGAAGATT	TATACATAAT	ACTCGCTTTG	GTTTGGTGTG	CGATGATTCG
	51	CGGATTATTC	GATCGGATTG	CGGCTGGGGG	TGGTTTGATT	ACGCTGCCTG
	101	CACCTCTTCT	GGCAGGTATT	CCTCCGTGTG	CGGCAATTGC	CACCAACAAG
	151	CTGCAAGCAG	CCGCTGCTAC	GTTTTCGGCT	ACGGTTTCTT	TTGCACGCCAA
	201	AGGTTTGATT	GATTGGAAGA	AAGGTCCTCC	GATTGCGCCA	GCATCGCTTG
	251	CAGGCGCGGT	GGTGCGTGCA	TTATCGGTCA	GCTTGGTTTC	CAAGATATT
25	301	CTGCTGCGGG	TCGTGCGGTT	TTTGTGTGTA	TTTGTGCGCG	TGATTTTGTG
	351	GTTTTGCGCC	AAGCTCGACG	CGAGTAAGGA	AGGCAAGGCC	AGAATGCTCT
	401	TTTTTCTGTT	CGGTCTGACG	GTTGCACACC	TTTTGCGGTT	TTACGACGGT
	451	GTGTTGCGAC	CGGCTGTCGG	CTCGTTTTTT	CTGATTGCCT	TTATTGTTTT
	501	GCTCGGCTGC	AAGCTGTGTA	ACGCGATGTC	TTACACCAAA	TTGCGCAAGC
30	551	TTGCTGCAAA	TCCTGTTTCG	CTATCGGTAT	TCCTGCTGCA	CGGTTGATTT
	601	ATTTTCCCGA	TTGCGGCAAC	GATGGCGGTC	GGTGCGTTTG	TCGGTGCAGAA
	651	TTTAGGTGCG	AGATTGTCGCG	TCCGCTTCGG	TCGGAAGCTG	ATTAAGCCCG
	701	TGCTGATTGT	CATCAGCATT	TCGATGGCTG	TGAATTTGTT	GATAGACGAG
	751	AGAAATCCGC	TGTATCAGAT	GATGTTTCG	ATGTTTTAA	

This encodes a protein having amino acid sequence <SEQ ID 436>:

	1	MEDLYIILALGLVAMIAGFI	DAIAGGGGLI	TLPALLLAGI	PPVSAIATNKI
	51	LQAAATFSA	TVSFARKGLI	DWKKGLPIAA	ASFAGGVVGA
	101	LLAVVPVLLI	FVALYFVFSP	KLDGSKGKGA	RMSFFLFGTL
	151	VFGPGVGSFF	LIAFIVLLGC	KLINAMSYTK	LANVACNLGS
40	201	IFPIAATMAV	GAFVGNLGA	RFVAVRGSKL	IKPELLIVISI
	251	RNPLYQMIVS	MF*		SMAVKLLIDE

ORF109a and ORF109-1 show 99.2% identity in 262 aa overlap:

		10	20	30	40	50	60
45	orf109a.pep	MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPFVSAIATNKLQAAATFSA					
	orf109-1	MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPFVSAIATNKLQAAATFSA					
		10	20	30	40	50	60
	orf109a.pep	TVSFARKGLIDWKKGLPIAAASFVGGVAGALSVSLVSKDILLAVVPVLLIFVALYFVFSP					
50	orf109-1	TVSFARKGLIDWKKGLPIAAASFVGGVAGALSVSLVSKDILLAVVPVLLIFVALYFVFSP					
		70	80	90	100	110	120
	orf109a.pep	KLDGSKGKARMSFFLFGTLVAPLLGFDYGVFGVGSFFLIAFIVLLGCKLLNAMS YTK					
55	orf109-1	KLDGSKGKARMSFFLFGTLVAPLLGFDYGVFGVGSFFLIAFIVLLGCKLLNAMS YTK					
		130	140	150	160	170	180
	orf109a.pep	LANVACNLGSLVFLFHGSIIFPIAATMAVGAFVGNLGAFAVRFGSKLIKPELLIVISI					
60	orf109-1	LANVACNLGSLVFLFHGSIIFPIAATMAVGAFVGNLGAFAVRFGSKLIKPELLIVISI					
		190	200	210	220	230	240
	orf109a.pep	LANVACNLGSLVFLFHGSIIFPIAATMAVGAFVGNLGAFAVRFGSKLIKPELLIVISI					
	orf109-1	LANVACNLGSLVFLFHGSIIFPIAATMAVGAFVGNLGAFAVRFGSKLIKPELLIVISI					

```

orf109-1      LANVACNLGSLSVFLHGSIIFFIAATMAVGAFVGVANLGARFAVFRFGSKLIKPLLVISII
              190      200      210      220      230      240
5 orf109a.pep  SMAVKLLIDERNPLYQMIVSMFX
              250      260
orf109-1      SMAVKLLIDERNPLYQMIVSMFX
              250      260

```

10 Homology with a predicted ORF from *N.gonorrhoeae*

ORF109 shows 98.3% identity over a 231aa overlap with a predicted ORF (ORF109.ng) from *N. gonorrhoeae*:

```

15 orf109.pep  MEDLYIILALGLVAMIAGFIDAAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA 60
orf109ng      MEDLYIILALGLVAMIAGFIDAAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA 60
20 orf109.pep  TVSFARKGLIDWKKGLPIAAASVFGGVAGALSVSLVSKDILLAVVPVLLIFVALYFVFSP 120
orf109ng      TVSFARKGLIDWKKGLPIAAASVFGGVAGALSVSLVSKDILLAVVPVLLIFVALYFVFSP 120
25 orf109.pep  KLDGSGKEGKARMSFFLFGTLTVATAFGLRRCVRTGCRLVFSDCLYCFARLQAVERDVLHQ 180
orf109ng      KLDGSGKEGKARMSFFLFGTLTVATAFGLRRCVRTGCRLVFSDCLYCFARLQAVERDVLHQ 180
30 orf109.pep  IGERCLQSWFAIGIPAARFDYFFDCGNDGRCVCRCEFRCEICRTLRFEA 231
orf109ng      IGERCLQSWFAIGIPAARFDYFFDCGNDGRCVCRCEFRCEICRTLRFEA 231

```

An ORF109ng nucleotide sequence <SEQ ID 437> was predicted to encode a protein having amino acid sequence <SEQ ID 438>:

```

30 1 MEDLYIILAL GLVAMIAGFI DAIAGGGGLI TLPALLLAGI PFVSAIATNK
51 LQAAAATFSA TVSFARKGLI DWKKGLPIAA ASFAGGVVGA LSVSLVSKDI
101 LLAVVPEVLLI FVALYFVFSP KLDGSGKEGA RMSEFFLGLT VATAFGLRL
151 CVRTGCRVLF SDCLYCFARL QAVERDVLHQ IGERCLQSWF AIGIPAARFD
201 YFPDCGNDGG RCVCRCEFR EICRPLRFEA D*

```

35 Further work revealed the following gonococcal DNA sequence <SEQ ID 439>:

```

40 1 ATGGAAGATT TATACATAAT ACTCGCTTTG GGTGTTGGTG CGATGATCGC
51 CGGATTATTAT GATGCGGATTG CGGGCGGGGG TGGTTTGATT ACGCTCGCTG
101 CACTCTTGTG GGCAGGTATT CTTCCCGTGT CGGCAATTGC CACCAACAAG
151 CTGCAAGCAG CCGCTGCTAC GTTTTCGGCT ACGGTTTCTT TTGCACGCAG
201 AGGTTTGATT GATTGGAAGA AAGGCTCTCC GATTGCCGCA GCATCGTTTG
251 CAGGCGGCGT GTCGCTGCA TTAICGTCAC GCTTGGTTTC CAAGATATT
301 TTGCTGGCGG TCGTGCCGGT TTTGTTGATA TTTGTCGCGC TGTATTTTGT
351 GTTTTCGCCC AAGCTCGACG GCAGTAAGGA AGGCAGAACC AGAATGTCTT
401 TTTTTCCTAT TTTTTCGAGC GTTGACCCGC TTTTGGGTTT TTACGACGCT
451 GTGTTCCGAC CGGGTGTGCG CTCGTTTATT CTGATTGCCT TTATTGTTTT
501 GCTCGGCTGC AAGCTGTGTA ACGGATGATC TTACACCAAA TTGGCGAAGC
551 TTGCTGTCAA TCTTGCTTGC CTATCGATCT TCCGTCTGCA CGGTTCCGAT
601 ATTTTCCGCA TTGTGGCAAC GATGCGGCTC GCTGCTTTG TCGSTCGCAA
651 TTTAGTGGCG AGATTGTGCG TCCGCTTCGG TTCGAAGCTG ATTAAGCCGC
701 TGCTGATTGT CATCAGCATT TCGATGCTGC TGAATTGTT GATAGACGAG
751 AGAAATCCGC TGTATCAGAT GATTGTTTCG ATGTTTATA

```

This corresponds to the amino acid sequence <SEQ ID 440; ORF109ng-1>:

```

55 1 MEDLYIILAL GLVAMIAGFI DAIAGGGGLI TLPALLLAGI PFVSAIATNK
51 LQAAAATFSA TVSFARKGLI DWKKGLPIAA ASFAGGVVGA LSVSLVSKDI
101 LLAVVPEVLLI FVALYFVFSP KLDGSGKEGA RMSEFFLGLT VAPLLFCGAT
151 VFGPGVGSFF LIAFIVLLCG KLINAMSYTK LANVACNLGS LSVFLHGSII
201 IFPIVATMAV GAFVGVANLGA RFAVRFGSKL IKPLLVIVISI SMAVKLLIDE
251 RNPLYQMIVS MF*

```

ORF109ng-1 and ORF109-1 show 98.9% identity in 262 aa overlap:

		10	20	30	40	50	60
5	orf109ng-1.pep	MEDLYIILALGLVAMIAGFIDAIAGGGGLITL	PALLAGIPPVSAIATNKLQAAATFSA				
	orf109-1	MEDLYIILALGLVAMIAGFIDAIAGGGGLITL	PALLAGIPPVSAIATNKLQAAATFSA				
		10	20	30	40	50	60
		70	80	90	100	110	120
10	orf109ng-1.pep	TVSFARKGLIDWKKGLPAAASFAGGVGALS	SVSLVSKDILLAVVPVLLIFVALYFVFSP				
	orf109-1	TVSFARKGLIDWKKGLPAAASFAGGVGALS	SVSLVSKDILLAVVPVLLIFVALYFVFSP				
		70	80	90	100	110	120
		130	140	150	160	170	180
15	orf109ng-1.pep	KLDGSKGKGARMSFFLFGTLVAPLLGFYD	GVFGVGSFFLIAFIVLLGCKLLNAMS	YTK			
	orf109-1	KLDGSKGKGARMSFFLFGTLVAPLLGFYD	GVFGVGSFFLIAFIVLLGCKLLNAMS	YTK			
		130	140	150	160	170	180
20		190	200	210	220	230	240
	orf109ng-1.pep	LANVACNLGSLSVFLLHGSIIFFPIVATMA	VGAFAFGANLGAFAVRFGSKLIKPLLI	VISI			
	orf109-1	LANVACNLGSLSVFLLHGSIIFFPIVATMA	VGAFAFGANLGAFAVRFGSKLIKPLLI	VISI			
		190	200	210	220	230	240
25		250	260				
	orf109ng-1.pep	SMAVKLLIDERNPLYQMIVSMFX					
	orf109-1	SMAVKLLIDERNPLYQMIVSMFX					
30		250	260				

In addition, ORF109ng-1 shows homology to a hypothetical *Pseudomonas* protein:

	sp P29942 YCB9_PSEUDE HYPOTHETICAL 27.4 KD PROTEIN IN COBO 3'REGION (ORF9)	
	>gi 94984 pir I 38164 hypothetical protein 9 - Pseudomonas sp >gi 551929	
	(M62866) ORF9 [Pseudomonas denitrificans] Length = 261	
35	Score = 175 bits (439), Expect = 3e-43	
	Identities = 83/214 (38%), Positives = 131/214 (60%), Gaps = 1/214 (0%)	
	Query: 41 PPVSAIATNKLQXXXXXXXXXXXXXKGLIDWKKGFLPIXXXXXXXXXXXXXXXXXXDKD 100	
	PP+ + TNKIQ R+G ++ K+ LP+ D+	
40	Sbjct: 43 PPLQTLGTNKLQGLFGSGSATLSYARRGHVNLEQLPMALMSAAGAVLGALLATIVPGDV 102	
	Query: 101 LLAVVPVLLIFVALYFVFSFKLDGSKGKGARMSFFLFGTLVAPLLGFYDGVFGVGSFF 160	
	L A++P LLI +ALYF P + G + +R++ F+F LT+ PL+GFYDGVFGPG GSFF	
45	Sbjct: 103 LKAILPELLIAIALYFGLKPNM-GDVQHSRVTFVFVFTLTLVPLIGFYDGVFGPGTGSFF 161	
	Query: 161 LIAFIVLLGCKLLNAMS YTKLANVACNLGSLSVFLLHGSIIFFPIVATMAVGAFAFGANLGA 220	
	++ F+ L G +L A ++TK N N+G+ VFL G+++ + M +G F+GA +G+	
	Sbjct: 162 MLGFVTIAGFVGLKATAHTKLFNFGSNVGAFAFGVLFFGAVLWKVGLMLGLGQFLGAQVGS 221	
50	Query: 221 RFAVRFGSKLIKPLLIIVISISMAVKLLIDERNPL 254	
	R+A+ G+K+IKPLL++SI++A++LL D +PL	
	Sbjct: 222 RYAMAKAKIILKLLVIVSIALAIRLLADPTHPL 255	

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 52

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 441>:

```

1  ..CTGCTAGGGT ATTGCATCGG TTATCGGTAC GGCTGTTGCA GCAAAACCCAG
5  51  CCGCAGACGG ATTATTTGGT CAAATTCCGA TCGTTTGGG CGAG. ATTTT
101  TGGTTTTCTG GGAATCTGATG ACGTCTATGC TCGGCATGG TTTGTGTTA
151  TCAATGATGTT TTTCGTGTT TCTACCAAGT TCGCCTGAT TCGCAATGTG
201  CCGCCGTCTC TGGCGGAAAT GAAGTCTTTT CGGGAAAGG TTAAGAAAA
251  ATCTCTGGCG GCGATCGGCC ATTCTTCGCT GTTGGATGTA AAAATTGGCG
10  301  CCGAGGTTGC CAAACGTTAT CTGCAAGTAC AAGGTTTTCA GGGGAAACCC
351  ATTAACCGTG AAGACGGGTC GGTTCGTATT GCCGCCAAA AAGGCACAAT
401  GAACAAATGG GGCATATATCT TTGCCCATGT TGCTTTGATT GTCATTTCGC
451  TGGCGGGGTT GATAGACAGT AACCTGCTGT TGAACTGGG TATGCTGACC
501  GGTCCGATTG TTCCGGACAA TCAGCGCGTT TATGCCAAGG ATTTT. AAGC
15  551  CCGAAAGTAT .TTTGGGTGC gTCCAATCTC TCATTAGGG GCAACGTCAA
601  TATTTCCG.A GGGGCAGAgT GCGGATGTGG TTTTCTGA

```

This corresponds to the amino acid sequence <SEQ ID 442; ORF110>:

```

1  ..LLGIASVIGT LLQNNQFQTD YLVKFSFWA XIFGFLGLYD VYASAWFVVI
51  MMFLVSTSL CLIRNVPPFW REMKSFREKV KEKSLAAMRH SSLLDVKRIAP
101  EVAKRYLEVQ GFQGKTIINRE DGSVLIAARK GTMNKNGYIF AHVALIVICL
151  GGLIDSNLLK LGLMLTGRIF RTIRRFMPRI XKPESXFCGV QSLI*GQRQY
201  FXRGRVRMWF S*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with ORF88a from *N.meningitidis* (strain A)

ORF110 shows 91.5% identity over a 188aa overlap with ORF88a from strain A of *N. meningitidis*:

```

25  orf88a.pep  10      20      30      40      50      60
      MSKSRRSFPLLSRPWF AFFSSMRFAVALLSLGLIASVIGT VLQNNQFQTD YLVKFSFWA
      orf110      |||||
      LLGLIASVIGT LLQNNQFQTD YLVKFSFWA
30  orf88a.pep  70      80      90      100     110     120
      QIFGFLGLYD VYASAWFVVI MMFLVSTSL CLIRNVPPFW REMKSFREKV KEKSLAAMRH
      orf110      |||||
      XIFGFLGLYD VYASAWFVVI MMFLVSTSL CLIRNVPPFW REMKSFREKV KEKSLAAMRH
35  orf88a.pep  130     140     150     160     170     180
      SSLLDVKIAP EVAKRYLEVQ GFQGKTIINRE DGSVLIAARK GTMNKNGYIF AHVALIVICL
      orf110      |||||
      SSLLDVKIAP EVAKRYLEVQ GFQGKTIINRE DGSVLIAARK GTMNKNGYIF AHVALIVICL
40  orf88a.pep  100     110     120     130     140     150
      GGLIDSNLLK LGLMLTGRIF RTIRRFMPRI XKPESXFCGV QSLIXGQRQY FXRGRVRMWF
      orf110      |||||
      GGLIDSNLLK LGLMLTGRIF RTIRRFMPRI XKPESXFCGV QSLIXGQRQY FXRGRVRMWF
45  orf88a.pep  160     170     180     190     200     210
      LNAADNGILVQDL PFEVKKKFKHIDFYNTGMPRDFASDIEVTDKATGEKLETRIRNVHPLT
      orf110      SX
50  orf88a.pep  250     260     270     280     290     300

```

However, ORF88 and ORF110 do not align, because they represent two different fragments of the same protein.

Homology with a predicted ORF from *N.gonorrhoeae*

ORF110 shows 88.6% identity over a 211aa overlap with a predicted ORF (ORF110.ng) from *N.gonorrhoeae*:

5	orf110.pep	LLGIASVIGTLLQQNQPTDYLVKGSFWA	30
	orf110.ng	MSKSRISPTLLSRFWAFFSSMRFAVALLSLLGIASVIGTVLQQNQPTDYLVKGPFWT	60
10	orf110.pep	XIFGFLGLYDVYASAWFVVMFLVYSTSLCLIRNVPPFWREMKSFREKVKESLAAMRH	90
	orf110.ng	RIFDFLGLYDVYASAWFVVMFLVYSTSLCLIRNVPPFWREMKSFREKVKESLAAMRH	120
15	orf110.pep	SSLLDVKIAPEVAKRYLEVQGFQGKTINREDGSLVIAAKGTMNKWGYIFAHVALIVICL	150
	orf110.ng	SSLLDVKIAPEVAKRYLEVQGFQGKTINREDGSLVIAAKGTMNKWGYIXAHVALIVICL	180
20	orf110.pep	GGLIDSNLLKLMGLTGRIFRTIRRFMPRIKXKPFSGVSLIKGQRYQFGRVVMWF	210
	orf110.ng	GRLINXNLLKLMGLAGSIFRNNRRVMFRIKSPESIWGGVSLIKGQRYQFGRVVMWF	240
25	orf110.pep	S 211	
	orf110.ng	S 241	

The complete length ORF110ng nucleotide sequence <SEQ ID 443> is predicted to encode a protein having amino acid sequence <SEQ ID 444>:

25	1	MSKSRISPTL	LSRFWFAFFS	SMRFAVALLS	LLGIASVIGT	VLQQNQPTD
	51	YLVKFGPFWT	RIFDFLGLYD	VYASAWFVVI	MMFLVYSTSL	CLIRNVPPFW
	101	REMKSFREKV	KEKSLAAMRH	SSLLDVKIAP	EVAKRYLEVR	GFQKQTVSRE
	151	DGSLVIAAKK	GTNMKWGYIX	AHVALIVICL	GRLINXNLL	KLMLAGSIF
	201	RNNRRVMFRI	SKPESIWGGV	QSLIKGQRYQ	FGRVVMWF	S*

30 Based on the putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 53

The following DNA sequence was identified in *N.meningitidis* <SEQ ID 445>:

35	1	ATGCCGCTCTG	AAACACGCCT	GCCGAACCTT	ATCCGCGTCT	TGATATTTGC
	51	CTCGGGTTTC	ATCTCTCTGA	ACGCTGTGTC	GGAACAAACC	GCGCAAAACG
	101	TTACCCCTGCA	AGGCGAAACG	ATGGGCGAGA	CCTATACCGT	CAATACCTTT
	151	TCAATAATCT	GGGACAAACT	CCCTCACCCT	GCGGAATATC	AAAAAGCGAT
	201	CGATGACGCG	CTTAAGAAGG	TCAACCGCGA	GATGTCACAC	TATCAGCCCG
40	251	ACTCCGAAAT	CAGCGGGTTC	AACCAACACA	CAGCCGGCAA	GCCCCCTCCG
	301	ATTTCAAGCG	ACTTCGCACA	CGTTACTGCC	GAAGCGCTCG	GCCTGAACCG
	351	CCTGACACAC	CGCGCGCTGG	ACGTAAACGT	CGGCCCTTTG	GTCAACCTTTT
	401	GGGGATTTCG	CCCCGACAAA	TCCGTTACCC	GTGAAACGCT	GCCGGAACAA
	451	ATCACAACAGG	CGGCATCTTA	TACGGGCATA	GACAAAATCA	TTTTGAACAA
45	501	AGGCAAAAGT	TACGCTTCCT	TGAGCAAAAC	CCACCCCAAG	GCCTATTTGG
	551	ATTTATCTTC	GATTGCCAAA	GGCTTCGGCG	TTGATAAAGT	TGCGGGCGAA
	601	CTGGAATAAT	ACGGCAATCT	AAATTATCTG	GTCCGAAATCG	GCGGCGAGTT
	651	GCACGCGCAA	GGCAAAACG	CGCGGGCGCA	ACCTGTGGCG	ATCGGTATCG
	701	AGCAGCCCAA	TATCGTCCAA	GCGCGGATTA	CGCAGTATAT	CTCTCCGCTG
50	751	AAACAACGTT	CGCTTGGCAC	TTCGCGGAT	TACCGTATTT	TCCAGTCGGA
	801	TAAAAACGCG	AAACGCCTCT	CCCATATCAT	CAACCCGAAC	AACAACACGA
	851	CCATCAGCCA	CAACCTCGCC	TCCATCAGCG	TGGTCGCGAG	CAGTGGGATG
	901	ACGGCGGACG	GTTGTGCCAC	AGGATTATTC	GTATTGGGCG	AAACCCGAAGC
	951	CTTAAGACTG	GCAGAGCGCG	AAAAATCGC	TGTTTTCCTG	ATTGTCAGGG

1001 ATAAAGGCGG CTACCGCACC GCCATGCTTT CCGAATTGA AAACTGCTC
1051 CGCTAA

This corresponds to the amino acid sequence <SEQ ID 446; ORF111>:

5 1 MPSETRLNPF IRVLIFALGF IFLNACSEQT AQTVTLQGET MGTITYVKYL
51 SNNRDKLESF AEIQKRIDDA LKEVNRQMST YQPDSEISRF NQHTAGKPLR
101 ISSDFAHVTA EAVRLNRLTH GALDVTVGPL VNLWGFSGDPK SVTREPSPEQ
151 IQQASYTGI DKIIKQKGD YASLSKTHPK AYLDLSSIAK GFGVDKVAEG
201 LEKYGIQNYL VEIGGELHGK GKNARGEFWR IGIEQPNIVQ GGNTQIIIVFL
10 251 NNRSLATSGD YRIFHVDKNG KRLSHIINFN NKRPISHNLA SISVVDASAM
301 TADGLSTGLF VLGETEALKL AEREKLAFLV IVRDKGGYRT AMSSEFEKLL
351 R*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF111 shows 96.9% identity over a 351aa overlap with an ORF (ORF111a) from strain A of *N.*

15 *meningitidis*:

		10	20	30	40	50	60
orfl11a.pep		MPSETRLNPFIRLTIFALSFIFLNACSEQT	AQTVTLQGETMGTITYVKYLSNNRDXLPSP				
orfl111		MPSETRLNPFIRVLIFALGFIFLNACSEQT	AQTVTLQGETMGTITYVKYLSNNRDKLPSP				
		10	20	30	40	50	60
		70	80	90	100	110	120
orfl11a.pep		AEIQKRIDDALKEVNRQMSTYQPDSEISRFN	QHTAGKPLRISDDFAHVTA	EAVHLNRLTH			
orfl111		AEIQKRIDDALKEVNRQMSTYQPDSEISRFN	QHTAGKPLRISDDFAHVTA	EAVRLNRLTH			
		70	80	90	100	110	120
		130	140	150	160	170	180
orfl11a.pep		GALDVTVGFLVNLWGFSGDPKSVTREPSPEQ	IKQAASVTGIDKIIKQKDYASLSKTHPK				
orfl111		GALDVTVGFLVNLWGFSGDPKSVTREPSPEQ	IKQAASVTGIDKIIKQKDYASLSKTHPK				
		130	140	150	160	170	180
		190	200	210	220	230	240
orfl11a.pep		AYLDLSSIAKGFGVDKVAEGLEKYGIQNYL	VEIGGELHGKKNARGEFWRIGIEQPNIVQ				
orfl111		AYLDLSSIAKGFGVDKVAEGLEKYGIQNYL	VEIGGELHGKKNARGEFWRIGIEQPNIVQ				
		190	200	210	220	230	240
		250	260	270	280	290	300
orfl11a.pep		GGNTQIIIVPLNNRSXATSGDYRIFHVDKSG	KRLSHIINFNNKRPISHNLA	SISVVDASAM			
orfl111		GGNTQIIIVPLNNRSXATSGDYRIFHVDKSG	KRLSHIINFNNKRPISHNLA	SISVVDASAM			
		250	260	270	280	290	300
		310	320	330	340	350	
orfl11a.pep		TADGXSTGLFVLGETEALKLAEREKLAFLV	IVRDKGGYRTAMSSEFEKLLRX				
orfl111		TADGLSTGLFVLGETEALKLAEREKLAFLV	IVRDKGGYRTAMSSEFEKLLRX				
		310	320	330	340	350	

The complete length ORF111a nucleotide sequence <SEQ ID 447> is:

55 1 ATGCCGCTCTG AAACACGCCT GCCGAACCTT ATCCGCACCT TGATATTTGC
51 CCTGAGTCTT ATCTTCCTGA ACGCCTGTTC GGAACAACCC GCGCAACCG
101 TTACCCCTGCA AGGTGAACCG ATGGGCACGA CCTATACCGT CAATACCTT
151 TCAAAATATC GGGACNACT CCCNTCACTT GCCGAAATAC AAAANCGCAT
201 CGATGACGCG CTTAAAGAAG TCAACCGGCA GATGTCACCC TATCAGCCCG
251 ACTCCGAATC CAGCCGCTTC AACCAACACA CAGCCGCGAA GCCCCTCCGC
301 ATTTCAAGC ACTTCGACA CTTACTGCG CAGCCGCTCC ACCTGAACCG
351 CCTGACACAC CGCGCGCTG GATTAACCG GTGACCTTG GTGACCTTG
60 401 GGGGATTCCG CCCCGACAAA TCCGTTACCC GTGAACCGTC GCGCGAACAA

5
10
1001
1051

```

451 ATCAAAACAG CAGCATCTTA TACGGGCATA GACAAAATCA TTTTGAACA
501 AGGCAAGAT TACGCTTCC TACGCAAAAC CCACCCCAAG GCCTATTATGG
551 ATTTATCTTC GATTGCCAAA GCCTTCGGCG TTGATNANGT TCGGGCGGAA
601 CTGGAATAAT ACGGCATTCA AAATTTATCTG GTCGAAATOG GCGGAGAGTT
651 CTACGGCAAA GNCAAAACG CGCGCGGCGA ACCTTGGCGC ATCGGCATCG
701 AACAGCCCAA CATCTGCCAA GCGCGCAATA CGCAGATTAT CGTCCCGCTG
751 AACAAACGTT GNTTGGCCAT TCCGGCGCAT TACCGTATTT TCCAGTCTGA
801 TAAAGCGCGC AAACGCTCT CCCATATCAT TAATCGGAAC AACAAACGAC
851 CCATCAGCCA CAACCTGCCT TCCATCAGCG GNTTCGAGA CAGTGGGATG
901 ACGGCGGACG GCTTNTCCAC AGGATTATTC GTATTGGCGC AAACCGAAGC
951 CTTAAGCTG GCAGAGCGCG AAAAAGCTGC TGTTTTCTGT ATTGTGAGG
1001 ATAAAGGCGG CTACCGCACC GCCATCTCT CGGAATTGA AAAAGTCTC
1051 CGCTAA

```

This encodes a protein having amino acid sequence <SEQ ID 448>:

15
20
351

```

1 MPSETRLPNF IRTLIFALSF IFLNACSEQT AQTVTLQGET MGTTYTVRYL
51 SNNRDXLSP AEIQXRIDDA LKEVNRQMS YQPDSEISR NQHTAGKPLR
101 ISSDFAHVTA EAVHINRLTH GALDVTGVL VNLWGFDPK SVTREPSPE
151 IKQAASTYGI DKILLKQKGD YASLSKTHP AYLDLSSIAK GFGVDKVAE
201 LEKYGIQNYL VETGGLGKX KXNARGEFWR IGTEQPNIVQ GNGTQIIVPL
251 NNRXATSGD YRIFHVDKSG KRLSHIINPN NKRPISHNLA SISVXADSAM
301 TADGXSTGLF VLGETEALKL AEREKLAFLV IVRDKGGYRT AMSSEFEKLL
351 R*

```

Homology with a predicted ORF from *N.gonorrhoeae*

25 ORF111 shows 96.6% identity over a 351aa overlap with a predicted ORF (ORF111.ng) from *N. gonorrhoeae*:

30
35
40
45
50
55
60

```

      10      20      30      40      50      60
orf111ng MPSETRLPNLIRALIFALGFIFLNACSEQTAQVTTLQGETMGTTYTVRYLSNNRDKLPS
      10      20      30      40      50      60
orf111 MPSETRLPNFIRVLIFALGFIFLNACSEQTAQVTTLQGETMGTTYTVRYLSNNRDKLPS

      70      80      90      100     110     120
orf111 AKIQKRIDDALKEVNRQMSYQTDSEISRNFQHTAGKPLRISSDFAHVTAEAVRINRLTH
      70      80      90      100     110     120
orf111 AEIQKRIDDALKEVNRQMSYQPDSEISRNFQHTAGKPLRISSDFAHVTAEAVRINRLTH

      130     140     150     160     170     180
orf111ng GALDVTGVLVNLWGFDPKSVTREPSPEIKQAASTYGDIIKLQCKDYASLSKTHPK
      130     140     150     160     170     180
orf111 GALDVTGVLVNLWGFDPKSVTREPSPEIKQAASTYGDIIKLQCKDYASLSKTHPK

      190     200     210     220     230     240
orf111ng AYLDLSSIAKFGVDKVAELEYKIQNYLVEIGGELHGKGNAGEFWERIGIEQPNIIQ
      190     200     210     220     230     240
orf111 AYLDLSSIAKFGVDKVAELEYKIQNYLVEIGGELHGKGNAGEFWERIGIEQPNIVQ

      250     260     270     280     290     300
orf111ng GGNTQIIVPLNNRSLATSGDYRIFHVDKNGKRLSHIINPNKRPISHNLAISVVDMSAM
      250     260     270     280     290     300
orf111 GGNTQIIVPLNNRSLATSGDYRIFHVDKNGKRLSHIINPNKRPISHNLAISVVDMSAM

      310     320     330     340     350
orf111ng TADGLSTGLFVLGETEALKLAEEKLAFLVIRVDKGGYRTAMSSEFAKLRLX
      310     320     330     340     350
orf111 TADGLSTGLFVLGETEALKLAEREKLAFLVIRVDKGGYRTAMSSEFAKLRLX

```

The complete length ORF111ng nucleotide sequence <SEQ ID 449> is:

1 ATGCCGCTCTG AAACACGCCT GCCGAACCTT ATCCGCGCCT TGATATTTGC
 51 CCTGGGTTTC ATCTTCCTGA ACGCCTGTTC GGaaCaaccG GCGCAaaccg
 101 TTACCTTGCA AGGCGAAacg aTGGGTACGA CCTATACCTT CAATACCTT
 151 TCRAATATAT GGGCAAACT CCCCTCCCTC GCCAAAATCA AAAAGCGCAT
 201 TGATGATGCG CTTAAGAAG CTCAACGGCA GATGTACACC TACCAGACCG
 251 ATTCGGAAT CAGCGGTTT AACCAACACA CAGCGCGCAA GCCCTCCGCG
 301 ATTTCAAGCG ATTTGCGACA CGTTACGCGC GAAGCGTGCC GCCTGAACCG
 351 CCTGACTCAC GGGCGACTGG ACGTAAACGT CGGCCCTTTG GTCAACCTTT
 401 GGGGGTTTCG CCGCGACAAA TCGGTTACCC GTGACCGTC GCCGGAACAA
 451 ATCAACAGG CGGCATCTTA TACGGGCATA GACAAAATCA TTTTGACACA
 501 AGGCAAGAT TACGCTTCTT TGAGCAAAAC CACCCCAAAA GCCTATTGTG
 551 ATTTATCTTC GATTGCCAAA GGCTTCGGGG TTGATAAAGT TCGCGGCGAA
 601 CTGGAATAAT ACGGCATTCA AAATTATCTG GTCGAAATcg gggcgGAGTT
 651 GCACGGCAAA GGCAAAATCT CGCACGGCGA ACCGTGGCGC ATCGGTATAG
 701 AGCAACCCAA TATCATCCAA GcgGCAata CGCAGATTAT cgtcccgctg
 751 aaCaaccgctt cgtCTGCCAC TTCGGCGGAT TACgcTATT tccagctcgA
 801 TAAAAAGggc aaacgccttt ccaacTCTAT CAATCCCAaC aaCAACgac
 851 cCATCAAGca caactctggc tcaatcagcg tggctcAGA GCGTCAATG
 901 ACGCGGACG GTTtatCCAC AGGATTATT GTTTTAGGCG AAACCGAAGC
 951 CTTAAGCGCTG CAGACAACAG AAAAATCTGC GTTTTTCTTA ATTTGTCGGG
 1001 ATAGGACGG CTACCGCAAC GCCATGTCTT CCGAATTTCG CAACTGCTC
 1051 CGATAA

This encodes a protein having amino acid sequence <SEQ ID 450>:

1 MPSETRLPNL IRLALFALGF IFLNACSEQT AQTVTLQGET MGTTTYVKYL
 51 SNNRDKLPSP AKIQKRIDDA LKEVNRQMS T YQTDSEISR F NQHTAGKPLR
 101 ISSDFAHVTA EAVRLNRLTH GALDVTGPL VNLNGGPGDK SVTPRESPEQ
 151 IKQAASTYGI DKILQQQGD YASLSTHPK AYLDLSSIAK GFGVDKVAGE
 201 LEKYGIQNYL VEIGSELHGK GKNNAHGEWR IGIEQPNIIQ GNGTQIIVPL
 251 NNRSLATSGD YRIFVHDKNG KRLSHLINPN NKRPISHNLA SISVSDSAM
 301 TADGLSTGLF VLGETEALRL AEQKLAFLVL IVRDKDGYRT AMSSEFAKLL
 351 R*

This protein shows homology with a hypothetical lipoprotein precursor from *H. influenzae*:

sp|P44550|YOJL_HAEIN_HYPOTHETICAL_LIPOPROTEIN_H10172_PRECURSOR >gi11074292|pir|4
 hypothetical_protein_H10172 - Haemophilus influenzae (strain Rd KW20)
 35 >gi11573128 (U32702) hypothetical [Haemophilus influenzae] Length = 346
 Score = 353 bits (896), Expect = 9e-97
 Identities = 181/344 (52%), Positives = 247/344 (71%), Gaps = 4/344 (1%)
 Query: 7 LPNLIRLALFALGFIFLNACSEQTACTVTLQGETMGTTTYVKYLSNNRDKLPSPAKIQKR 66
 + LI +I + L AC ++T ++L G+TMGTTY VKYL + S K +
 40 Sbjct: 1 MKKLISGIIAVAMALSLAACQKET-KVISLSGKTGMTTYHVKYLDGGSITATSE-KTHEE 58
 Query: 67 IDDALKEVNRQMSYQTDSEISRFOHT-AGKPLRISSDFAHVTA EAVRLNRLTHGALDV 125
 I+ LK+VN +MSTY+ DSE+SRFNQ+T P+ IS+DFA V AER+RLN+VT GALDV
 45 Sbjct: 59 IEALKLVNARMSTYKRDSELSRFNQTQVNTPIEISADFKAFLVLAERILNKVTEGALDV 118
 Query: 126 TVGPLVNLWGFGPKDSVTPRESPEQIKQAASVTGIDKILQQQGDYASLSKTHPKAYLDL 185
 VGP+VNLWGFGPKL + ++P+EGK L K+ R+LSK P+ Y+DL
 50 Sbjct: 119 TVGPPVNLWGFGPEKRPEKQPTFEQLAERQAWVGIDKITTLDTNKEKATSKALPQVYVDL 178
 Query: 186 SSIAGKFGVDKVAGELEKYGIQNYLVEIGSELHGKGNNAHGEWRIGIEQPNIIQGGNTQ 245
 SSIAGKFGVD+VA +LE+ QNY+VEIGGE+ KKKN G+PW+I IE+P +
 55 Sbjct: 179 SSIAGKFGVDQVAEKLEQLNAQNYMVEIGGEIRAKGKNIEGKFWQIAIEKPTTTGERAVE 238
 Query: 246 IIVPLNNRSLATSGDYRIFHVDKNGKRLSHLINPNKRPISHNLA SISVSDSAMTADGL 305
 ++ LNN +A+SGDYRI+ ++NGKR +H I+P PI H+LAST+V++ ++MTADGL
 Sbjct: 239 AVIGLNNMGMASSGDYRI-PEENGKRAFEIDPKTGYPYIQHHLASTTVIAPTSMTADGL 297
 60 Query: 306 STGLFVLGETEALRLAEQKLAFLVLRDKDGYRTAMSSEFAKL 349
 STGLFVLGE +AL +AE+ LAV+LI+R +G+T S S F KL
 Sbjct: 298 STGLFVLGEDKALEVAERNLAVYLIIRTDNGFVTKSSSAFFKL 341

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

orf35a	QRLAIPAEAEVLYAQQAAYAANTLFLGLRAADRGDDVYAADPSRQKLWLRFIGGRSHQNI	310	320	330	340	350	360
5	orf35.pep	40	50	60	70	80	90
	orf35a	GAAADGWRKGVQIGGEVFEVRQNEGSXLAIGVMGGRAGQHASVNGKGAAGSDLYGYGGGV					
	orf35a	370	380	390	400	410	420
10	orf35.pep	100	110	120	130	140	150
	orf35a	430	440	450	460	470	480
15	orf35.pep	160	170	180	190	200	210
	orf35a	490	500	510	520	530	540
20	orf35.pep	220	230	240	250	260	
	orf35a	550	560	570	580	590	600
25	orf35a	KEAALS LKWLFX	610	620			

The complete length ORF35a nucleotide sequence <SEQ ID 453> is:

1	ATGTTCCAG	CTCAGCTGG	TTCAAATACT	CGTTCTACCA	AAATCGGGCA
5	CGATGCCGAT	TTTTCATITT	CAGACAAGCC	GAAACCCGGC	ACTTCCCATT
10	ATTTTCCAG	CGGTAAACC	GATCAAATT	CATCCGAATA	TGGGTATGAC
15	GAAATCAATA	TCCAAGGTAA	AAACTACAA	AGCGGCATAC	TGCGCTCGCA
20	TAATATGCCC	GTGTTAAGA	AATATATTAC	AGATACTTAC	GGGATAATT
25	TAAAGGATGC	GTTTAAAGAG	CAATTACAGG	ATTATACAAA	AACAAGACCC
30	GAAAGCTTCG	AGAGAAATA	AAACCGGCT	GAGGAGGCT	ATATAGACAA
35	GCTTTGGACCA	AAATTTAGTA	TACTCAACAA	GAAACACCC	GATTTAATTA
40	ATAAATTGGT	AGAGATTCC	GTACTCACT	CTCATGTAA	TACATCAAG
45	ACTAGTCTCA	ACACATCTT	CAATAAAAA	TTACACGTCA	AAATCGAAAA
50	CAAAATCCAC	CTGCGCGGAC	AGGTGTTGGA	ACTGACCAG	ATGACGCTGA
55	AAGATTCCCT	TTGGGAACCG	CGCGCCATT	CCGACATCCA	TATGCTGGAA
60	ACTTCCGATA	ATGCGCGCAT	CGCCTGAAC	ACGAAGATG	AAAACTGAC
65	CGTCCATAAA	CGGTATCAGG	CGCGTCGGA	TTTCCTGTT	GGCTACGACG
70	TGCGGGAGTC	GGACAAACCC	CGCCTGACCT	TTGAAGAAAA	AGTCAGCGCA
75	CAATCCGCGC	TGGTTTGGGA	ACGCGGCGC	GAAATCTGA	AAACGCTCGA
80	CGGCGCAGAA	CTGATTGCGG	CGGAAGAGC	AGACTCTAAT	TCGTTTGGCT
85	TTAAACAAAA	TTACCGGCAG	GGACTGTACG	AATTAITGCT	CAAGCAATGC
90	GAGGCGGGAT	TTTGCTTGGG	CGTGCAGCGT	TTGGCTATCC	CCGAGGCGGC
95	AGCGGTTTTA	TATGCCCAAC	AGGCTTATGC	GGCAATACT	TTGTTTCGGC
100	TGCGTCCGCG	CGACAGCTGT	ATGCTCCGCA	TCCCTCCCGT	
105	CAAAATTCAT	GCTGCGGCTT	CATCGGCGCT	CGCTCGATAC	AAATATATCA
110	GGCGGCGCG	GCTGCGGACG	GGCGGCGCAA	ACGGGTGCAA	ATCGGCGCGC
115	AGGTGTTTGT	ACGGCAAAAT	GAAGCGACGC	GGCTGCCAAT	CGCGGTGATG
120	GGCGGCGAGG	CTGCGCCAGCA	CGCATCAGTC	AACGGCAAG	CGGCTCGCGC
125	AGCGAGTTAT	TGCGATGGTT	ATGGCGGGGG	TGTTTATGCT	CGGTGCAATC
130	AGTTGCGCGA	TAAACAAACG	GGTGGGTATT	TGGACGGCTG	GTTCGATATC
135	CAACGTTTCA	AACACCGCAT	CAATGATGAA	AACCGTGGCG	AACGCTACAA
140	AACCAAAAGT	TGGACGGCTT	CTGTGGAAGG	CGGCTACAA	CGGCTTGTGG
145	CGGAAGCGGT	TGTCGGAATA	GGCAATAATG	CCTGCAACCG	
150	CAGGCGCAGT	TTACCTACTT	GGCGTAAAC	GGCGGCTTTA	CCGACAGCGA
155	GGGACGCGCG	CTCGGACTCG	TGCGCAGCGG	TCAGTGGCAA	AGCGCGCGCG
160	GCATTGCGGC	AAAAACCGGT	TTTGCTTTGC	GTAACGGTGT	CATCTTCGTC
165	CTTTTTCGCG	CTTTTAATGT	TTTGCACAGG	TCAAAATCTT	TCGGCTGTGA
170	AAATGACGCG	GAAACACATA	CGCTGCGAGG	CAGGCGCGCG	CTCGAAGGCG
175	GTTTGGCAT	TGCAAGCGCT	ATGATGTCGC	ATATGTCGCG	ACGCATCCGA
180	TACGCAAAA	GGACGACGCG	CGACAAAGAA	GCGCATTTGT	CGCTCAATAT
185	GCTGTTTGA				

This encodes a protein having amino acid sequence <SEQ ID 454>:

```

1  MFRAQLGSNT RSTKIGDDAD FSFSDKPKPG TSHYFSSGKT QDNSSEYGYD
51  EINIQGNKYN SGILAVDNMP VVKYITDTY GDNLKDAVKK QLQDLYKTRP
101 EAWEEKNKRT EEAYIEQLGP KFSILKQKNP DLINKLVEDS VLTPHSNTSQ
151 TSLNNIPNKK LHVKIENKSH VAGQVLELTK MTLKDLWEP RRHSDIHMLE
201 TSDNARIIRLN TKDEKLTVHK AYQGGADFLF GYDVRESDKP ALTFEKVSG
251 QSGVVLERRP ENLKTLDGRK LIAAEKADSN SEAFKQNYRQ GLYELLKQC
301 EGGFCVLGVQR LAIPEAEAVL YAQQAYANT LFGRAADRQ DDVYAADPSR
351 QKLWLRFIGG RSHQNTIRGA AADGRKGVQ IGGEVFVRQN EGSRLAIGVM
10 401 GGRAGQHASY NGKGGAGAGS YLHGCGGVYA AWHQLRDKQT GAYLDSWLQY
501 QAOFTYLGVN GGFPTDSEGT VGLGSGGWQ SRAGIRAKTR FALRNVGNLQ
551 PFAAFNVLHR SKSPGVEMDG EKQTLACRTA LEGRFGIEAG WKGHMSARIG
601 YGKRTDGDKE AALSLEWLF*

```

15 Homology with a predicted ORF from *N.gonorrhoeae*

ORF35 shows 51.7% identity over a 261aa overlap with a predicted ORF (ORF35ngh) from *N.*

gonorrhoeae:

```

      orf35.pep                                PCRRQGGDDVYAAHASRQKLWLRFIGGRSHQNIIRG      34
      orf35ngh                                FTKVQERDDIIAYQAQAANTLFLARLNDKNSIDFDRTLFRKGLWLRVLDGHSNQWVQG      370
20  orf35.pep                                GAA-ADGWRKGVQIGGEVFRQNEGSXLAIGVMGRAGQHASVNGKG--GAAGSDLYGYG      91
      orf35ngh                                KTAPEVGEYRKGVQLGGEVFTWQNESNQLSIGLGGQAEQRSTFRNPDTNLTITGNVKGFG      430
25  orf35.pep                                GGYYAANHQLRDKQTGAYLDGWLYQRFKHRINDENRAERYKTKGWTASeGGVGNALVAE      151
      orf35ngh                                AGVYATWHQLDKQKTGAYVDSWMQYQRFHRINTEYATERFTSKGITASIEAGYNALLAE      490
30  orf35.pep                                GIVGKGNNVRFYLPQQAQFTYLGVNGGFTDSEGTAVGLLGSQGWQSRAGIRAKTRFALRN      211
      orf35ngh                                HFTKKGNSLRVYLPQQAQLTYLGNGKFSDSENAQVNLGLSRQLQSRVGVQAKAQFAFTN      550
35  orf35.pep                                GVNLPQFAAFNVLHRSKSPGVEMDGKQTLAGRTALEGRFGIEAGWKGHMSA      263
      orf35ngh                                GVTFQPFVAVNSIYQKQKPFGEIDGRRVINNKTVIETQLGVAARIKSHLTLOASFNRQT      610

```

A partial ORF35ngh nucleotide sequence <SEQ ID 455> is predicted to encode a protein having

partial amino acid sequence <SEQ ID 456>:

```

40 1  .KKLADRNSEY KKEETYHIKS NGFTYNIPa LFPKHPPDFP ENINNSKKIS
51  FVDKEYTEDY LVGFARGEGV EKRNGEEKFP LRQYFKDCVN TENSNNDNCK
101 ISSFGNYGPI LKSDIFALA SQIKNSHINS ELLSVNGYIE WLRPTNLKLT
151 GWQEHLYAGL DPFHYIEVTD NSHVIGQITD LGAELITNSL WKPRWNSNID
201 YLITKNAEIR FNTKNESLVL KEDYAGGARF RFAYDLKDKV PEIPVLTFEK
251 NITGTSDIIF EGKALNDLKH LDGHQIVKVN DTAOKDAFRL SSKYRKGIYT
45 301 LSLQQRPEGF FTKVQERDDI AIYAQQAQAA NTLFALRLND KNSDIFDRTL
351 PRKGLWLRVI DGHSNQWVQG KTAPEVGEYRK GVQLGGEVFT WQNESNQLSI
401 GLMGGAQEQR STFRNPDTDN LTTGNVKGFG AGVYATWHQL QDKQGTAYVD
451 SWMQYQFRH RINTYATER FTSKGITASI EAGYNALLAE HFTKKGNSLR
501 VYLQPAQLT YLVNGKFS D SENAQVNLGL SRQLQSRVGV QAKAQFAFTN
50 551 GVTFFQFVAV NSIYQKQKPF VEIDGDRVI NNTKTVIETQL GVAARIKSHL
601 LTQASFNRQT SKHHHAKQGA LNLQWTF*

```

Based on this prediction, these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 55

55 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 457>:

1 ..CGCGAATATG TTCAGTCTC TATAGATTG TTCAGTGTGG GTAAATCGGG
 51 GGGCGGTATA CCTAAGCTA AGCCTGCTTT TGATGCGAAA CCGAGATGGG
 101 AGTTTCATAG GAACCTTAAT AAATTGACAA CTGCTGAGCA GGTGGAGAAA
 151 AATCTTCAGG AAACAGAAAG AAGGAGTCAG AGTAGTCAGT TTAAAGCCCA
 5 TCGCGCAACGA GAATGGGAAA ATAAACACGG GTTAGATTTT AATCATTTTA
 201 TAGGTGCTGA TATCAATAAA AAAGCGACAG TAACAGGAGG GCATAGTCTA
 251 ACCCTGGTGG ATGTACGGGT GATACACAA ACCTCGGCAC CTGATAAACA
 301 TGGGGT.TTA TCAAGCGACA GTGGAAATTN A

This corresponds to the amino acid sequence <SEQ ID 458; ORF46>:

10 1 ..AEYVQFSIDL FSVGKSGGGI PKAKPVFDAK PRWEVDRLN KLTTRQVEK
 51 NVQETRRRSQ SSQFKAAQR EWENKTGLDF NHFIGGDINK KGTVGGHSL
 101 TRGDVRIQQ TSAFDKHGXL SSSDGNX

Further work revealed further partial nucleotide sequence <SEQ ID 459>:

15 1 ..GCAGTGTGCC TnCCGATGCA TGCACACGCC TCAaATTGCG CAAACGATTC
 51 TTTTATCCGG CAGSTTCTCG ACCGTGACGA TTTCGAACCC GACGGGAAAT
 101 ACCACCTATT CGGCAGCAGG GGGGAACTTG CCGAGCGCCA GTCTCATATC
 151 GGATTTGGGA AAATACAAAG CCATCAGTTG GGCAACCTGA TGATTCAACA
 201 GCGCGCATT AAAGGAAATA TCGGCTACAT GTCCGCTTT TCGGATCAG
 251 GGCACGAAGT CATTCCCCS TTGACAACAC ATGCCCTACA TTCGATTCT
 201 GATGAAGCCG GTAGTCCCGT TGACGGATT AGCCTTACC GCATCCATT
 351 GGACGATAC GAACGATC CCGCAGCAGG CTATGACAGG CCACGAGGG
 401 CGCGCTATCC CCCTCCCAA GCGCGAGCG ATATATACAG TACGACATA
 451 AAAGCGCTTG CCAAAATAT CCGCCTCAAC CTGACGACGA ACCGCAGCG
 501 CGGACAACGG CTTCGCCGACC GTTTCACAA TGCCGCTAGT ATGCTGACCG
 25 551 AAGGAGTAGG CGACGGATTC AAACGCGCCA CCGGATACAG CCGCGAGCTG
 601 GACAGATCGG GCAATGCCG CGAAGCCTTC AACGGCACTG CAGATATCGT
 651 TAAAACATC ATCGGGGCTG CAGGAGAAAT TGT

This corresponds to the amino acid sequence <SEQ ID 460; ORF46-1>:

30 1 ..AVCLPMHAHA SXLANDSFIR QVLDQRHFEP DGKYHLFGSR GELAEQRSHI
 51 GLGKIQSHQL NLMITQAAI KGNIGYIVRF SDHGHEVHSP FDNHSHSDS
 101 DEAGSPVDGF SLYRIHWG DY EHHPADGYDG PQGGGYPAPK GARDIYSYDI
 151 KGVAQNIRLN LTDNRSTGQR LADRFHNAGS MLTQGVGDGF KRATRYSPEL
 201 DRSGNAEAF NGTADIVKNI IGAAGEI

Computer analysis of this amino acid sequence gave the following results:

35 Homology with a predicted ORF from *N.gonorrhoeae*

ORF46 shows 98.2% identity over a 111aa overlap with a predicted ORF (ORF46ng) from *N. gonorrhoeae*:

orf46.pep	AEYVQFSIDLF SVGKSGGGI PKAKPVFDAK PRWEVDRLNKL TTR	45
orf46ng	PKTGVPFDGKGF NFEKHVKYD TKLDIQELSGGGI PKAKPVFDAK PRWEVDRLNKL TTR	217
orf46.pep	EQVEKNVQETRRRSQSSQFKAAHQREWENKTGLDFNHFIGGDKKGTVTGGHSLTRGDV	105
orf46ng	EQVEKNVQETRRRSQSSQFKAAHQREWENKTGLDFNHFIGGDKKGAVTGGHSLTRGDV	277
orf46.pep	RVIQQTSA PDKHGKLVSSDSGN	126
orf46ng	RVIQQTSA PDKHGKLVSSDSGN	298

A partial ORF46ng nucleotide sequence <SEQ ID 461> is predicted to encode a protein having partial amino acid sequence <SEQ ID 462>:

1 ..RLKHCHCHAR LGSFHRKQD GAHQRFGRYG ATQRLCRSSH PRLGSPKPCQ
 51 RTRHRSRQOY LYGSHPHORD WSCPGLKQLG RHHGTS CRAV ADXRDRICER
 101 EIRRQRQXCR CRILGKI PSLS IPKYPLKLEQ RYKKNITSS TVPPSNGKNV
 151 KLADQRHFKT GVFPDGGKGF NFEKHVKYD KLDIQELSGG GTPKAKPVFD

201 AKPRWEVDK LNKLT'TREQV EKNVQETRRR SQSSQKKAHA QREWEKNTGL
251 DFNHFIGGDI NKKGAVTGGH SLT^{RG}DVVRVI QOTSAPDKHG VLSSDSGN*

Further work revealed the complete gonococcal DNA sequence <SEQ ID 463>:

5 1 TTGGGCATTT CCGCRAAAT ATCCCTTATT CTGTCCATAC TGGCAGTGTG
10 51 COTGCCGATG CATGCACACG COTCAGATTT GGcaAACGAT CCTTTATTC
101 GgCaggttct CGaccCTCAG CATTTGCaac ccgacgGAA ATACCCACTA
151 TTcggCaGCA GGGGGAGCT TgcnagcGC aacggcATa tccgattggG
201 aaacaTAcas Agccatcagt Tgggccacct gatgatccaa caggcgccgG
251 ttgaaggaaA TAtcgGctac attgtccgct tttccgatca cgggcacaaa
301 ttcatttcgc ccttcGAcas ccaTGCCCTCA CATTCOGATT CTGACGAAGC
351 CGGTAGTCCC GTTGACGGAT TCAGCCTTTA CCGCATCCAT TGGACGGAT
401 ACGAACCCA TCCGCGCAG GGCTATGAGC GGCCACAGGG CGGCGCTAT
451 CCGGCTCCCA AAGGCGGAG GGATATATAC AGCTACGACA TAAAGGCGT
501 TGCCCAAAAT ATCCGCTCTA ACCTGACCGA CAACCGCAGC ACCGACACAC
551 GGCTTGCCGA CCGTTTCCAC AATGCCGGGG CTATGCTGAC GCAAGAGTGA
601 GGGACGGAT TCAAACCGCG CACCGGATAC AGCCCGGAGC TGGACAGATC
651 GGGCAATGCC gccGAAGCCT TCAACGGCAC TGCAGATATC GTCAAAACAA
701 TCATCGGGCG GGCAGAGAA ATGTGCGCGC CAGCGGATCG CCGCagggT
751 ATAAAGCGAAG GCTCAACACT TCCCTG CATG CACGGCTTGG GTCTGCTTC
801 CACCGAATAA AAGATCGCGC GCATCAACGA TTTGGCAGAT ATGGCGCAAC
851 TCAAGACTA TGCCGACGA GCCATCCGCG ATTTGGCAGT CCMAAACCCC
901 ATGCGCGCAC AAGGCATAGA AGCCCTGACG AATATCTTTA TGGCAGCAT
951 CCCCATCAA GGGATTGGAG CTGTCCGGGG AAAATACGGC TTGGGCGGCA
1001 TCACGGCACA TCCTGTCAAG CGGTGCGAGA TGGGCGCAT CGCATTTCCG
1051 AAAGGGAAT CCAGCGCTAG CGACAATTTT GCGGATCGGG CATACGCCAA
1101 ATACCCGTCC CTTTACCATT CCGCAATAT CGCPTCAAC TTGAGCAGCG
1151 GTTACGGCAA AGAAAAATC ACCTCTCTCA CCGTGCCGCC GTCAACCGCG
1201 AAAAATGTCA AACTTGGCAGA CCAACGCCAC CGAAGACAG CGCTACCGTT
1251 TGACGGTAAA GGGTTTCCGA ATTTTGAGAA GCACGTGAAA TATGATACGA
1301 AGCTCGATAT TCAAGAATTA TCGGGGGGGC GTATACCTAA GGCTAAGCCT
1351 GTGTTTGATG CGAAACCGAG ATGGGAGGTT GATAGGAGC TTAATAATT
1401 GACAACCTGT GAGCAGGTGG AGAAAAATGT TCAGGAACAG AGAAGAGGA
1451 CTCAGAGTAG TCAGTTTAAA GCCCATCGCG AACGAGAATG GGAAAAATAA
1501 AGCAGGTTAG ATTTTATATA TTTTATAGGT GGTGATATCA ATAGAAAGG
1551 CACAGTACA GAGGCGGATA GTCTAACCG TGGTGATGTA CGGCTGATAC
1601 AACCAACCTC GGCACCTGAT AACATGGGG TTTATCAAG CACAGTGGAA
1651 ATTAAAAAGC CTGATGGAAG TTGGGAGGTG AAAACGAAAA AAGGTGGGA
1701 AGTGATGACC AAGCACCCA TGTTCCTAAA AGATTGGGAT GAGGTAGAA
1751 TTAGGCGTGA AGTTACTTGG GCTTGGGAAA TAGAGTAAAT GCTTAGGAT
1801 AATAAATGCC AGGGTACAA TAAATCGGT ATTAATAATG AAGGATTATC
1851 CGAACCTAAT AGAACAGCAT ATCCCATTTA TGAATAG

This corresponds to the amino acid sequence <SEQ ID 464; ORF46ng-1>:

45 1 LGISRKISLI L SILAVCLPM HAHASDLAND PFIRQVLDRO HFEPDGKYHL
51 FGSRGELAXR NGHILGLNIQ SHQLGLHLMIO QAAVEGNIGY IVRFSDDHGH
101 FHPSPDNHAS HSDSDEAGSP VDGFSLYRIH WOGYEHHPAD GYDGPQGGEY
151 PAPKGARDIY SYDIKGVAQN IRLNLTDNRS TGQRLLDREHF NAGAMLTQGV
201 GDGFKRATRY SPFLDRSGNA EAFNGTADI VKNIIGAAGE IVGAGDAVQV
251 ISEGSNIAYM HGLGLLSTEN KMARINDLAD MAQLKDYAAA AIRDWAVQNP
301 NAAQGLERVS NIFMAAIFIK GIGAVRGKIG LGGTTHAFVK RSQMGLALIP
351 KGSRAVSNIF ADAYAKVPS PYSRRIASN LEQRYKQENI TSSTVPSNLP
401 KNVKLADQRH PKTGVPFDKG GFPMFSKHVK YDTKLDIQEL SGGCIKPAKE
451 VFDAKPRWEV DRKLNKLTTR QQVKNVQST RRRQSSEQFK AHQREWEKX
501 TGLDFNHFIG GDINKKGTVT GHGSLTRGDV RVIQOTSAPD KHGVYQATVE
551 IKKPDGSWEV KTKKGGKVM T KHTMFPKDW EARIAREVTS AWESRIMLKD
55 601 NKWQGTSGSG IKIEGFTEPN RTAYPIYE*

ORF46ng-1 and ORF46-1 show 94.7% identity in 227 aa overlap:

10 20 30 40
or46-1.pep AVCLPMHAHAXLANDSFIRQVLDROHFEPDGKYHLFGSRGELAXR
|||||
60 or46ng-1 LGISRKISLILSILAVCLPMHAHAXLANDSFIRQVLDROHFEPDGKYHLFGSRGELAXR
10 20 30 40 50 60
50 60 70 80 90 100

	orf46-1.pep	QSHGLGKIQSHQLGNLMIQQAATKNGIYIVRFSDDHGHEVHSPFDNHHASHSDSDEAGSP
	orf46ng-1	NGHIGLGNIQSHQLGHLMIQQAAVEGNIYIVRFSDDHGKHFHSPFDNHHASHSDSDEAGSP
5		70 80 90 100 110 120
	orf46-1.pep	110 120 130 140 150 160
	orf46ng-1	130 140 150 160 170 180
10		
	orf46-1.pep	170 180 190 200 210 220
	orf46ng-1	190 200 210 220 230 240
15		
	orf46-1.pep	I
	orf46ng-1	IVGAGDAVGIGISEGNSIAVMHGLLLSTENKMARINDLADMAQLKDYAAAAIRDWAVQNP
20		250 260 270 280 290 300

Homology with a predicted ORF from *N.meningitidis* (strain A)

25 ORF46ng-1 shows 87.4% identity over a 486aa overlap with an ORF (ORF46a) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
	orf46a.pep	LGISRKISLLSILAVCLPMHAHASDLANDSPTRQVLDQHFEPDGKHYHFGSRGELAEPR					
30	orf46ng-1	LGISRKISLLSILAVCLPMHAHASDLANDPFTIRQVLDQHFEPDGKHYHFGSRGELAKR					
		70	80	90	100	110	120
35	orf46a.pep	SGHIGLGNIQSHQLGNLFQQAQKNGTGYIVRFSDHGHEVHSPFFNHASHSDSDEAGSP					
	orf46ng-1	NGHIGLGNIQSHQLGHMLTQQAQVEGNTGYIVRFSDHGKHKFSPPFNHASHSDSDEAGSP					
		70	80	90	100	110	120
40	orf46a.pep	VDGFSLYRIHWDGYEHHPADGYDGGPGGGYPAKPGARDIYSYDIKGVAQNIRLNLTNRKS					
	orf46ng-1	VDGFSLYRIHWDGYEHHPADGYDGGPGGGYPAKPGARDIYSYDIKGVAQNIRLNLTNRKS					
		130	140	150	160	170	180
45	orf46a.pep	TGQRLVDRFHNFTGSMGLTQGVGDGFKRATRYSPELDRSGNAAEAFNGTADIVKNIIGAAGE					
	orf46ng-1	TGQRLADRFHNAGAMLTQGVGDGFKRATRYSPELDRSGNAAEAFNGTADIVKNIIGAAGE					
		190	200	210	220	230	240
50	orf46a.pep	IVGAGDAVQGISSEGSNIIVMHGLGLLSTENKMARINDLADMAQLKDYAAAAIRDWAVQNP					
	orf46ng-1	IVGAGDAVQGISSEGSNIIVMHGLGLLSTENKMARINDLADMAQLKDYAAAAIRDWAVQNP					
		250	260	270	280	290	300
55	orf46a.pep	NAAQGIEAVSNIFTAVIPVKYGIGAVRGKYGLGGITAHVPKRSQMGELALPKGSAVSDNF					
	orf46ng-1	NAAQGIEAVSNIFMAAIPKYGIGAVRGKYGLGGITAHVPKRSQMGALIALPKGSAVSDNF					
		310	320	330	340	350	360
60	orf46a.pep	ADAAYAKYPSPYHSNRIRNLEORYGKENTISTSTVPPSNGKNVKLANIKRHPKTKVPDPGK					
	orf46ng-1	ADAAYAKYPSPYHSNRIRNLEORYGKENTISTSTVPPSNGKNVKIADORHPKTKVPTDCK					

		370	380	390	400	410	420		
5	orf46a.pep	430	440	450	460	470			
		GFPNFEKDKVYDTRINTAVPQVN---PIDEVFVN--PKGSVGSAAHSWSITARIQYAKLP							
			:::	:::	:::	:	:	::: :	
	orf46ng-1	GFPNFEKHVYDKLD--IQELSGGGIPKAKPVFDAKPRWEVDKRLN-KLTITREQVEKVN							
		430	440	450	460	470			
10	orf46a.pep	480	490	500	510	520	530		
		RQGRTRIYPKPNKYSAPLEKPGPNNGYLDKFGNWTGKPSRTKQGFWEVDQLSKTGREQ							
		:::							
	orf46ng-1	QETRRRSQSQFKAHAQREWENTGLDTNFIHFGGDINKKGTVTSGHSLTRGVVRIQQTS							
		480	490	500	510	520	530		

The complete length ORF46a DNA sequence <SEQ ID 465> is:

15	1	TTGGCGCATTT	CCCGCAAAAAT	ATCCCTTTATT	CGTGCCACAT	TGGCAGTGTG
	51	CTCCGCGCATG	CATCCACACG	CTCCAGATT	CGGCAACATG	CTTTTTCTTC
	101	GCGACGATCT	CGACCGCTCAG	CTATTCGAAC	CGCACGGGAA	ATACCACCTTA
	151	TTCCGGCAGCA	GGGGGGAATC	TGCGCAGGCG	AGCGCTCGTA	TCGGAATTGG
20	201	AAACATACAAA	AGGCATTCAGT	TGGGCAACTC	GTTCATCCAG	CAGGCGGCCA
	251	TAAAGGAAA	TATCCGCTAT	ATTTCGCGT	TTTCCGATCA	CGGCGACGCA
	301	TGCCCATTC	CTTCGACAA	CCAGTCCTCA	CATTCCGATT	CTGATGAAGC
	351	CGGTAGTC	GTTCAGCGAT	TCCAGCTTTA	CGCATCCAT	TGGACGCGAT
	401	ACGACACCA	CTCCGCGCAT	GGCATATGAC	CGCCACAGGG	CGGCGGCTAT
	451	CCCGCTCCCA	AAGGCGGAG	GGGTATATAC	AGCTACGACA	TAAAGAGCGT
25	501	TGCCCAAAAT	ATCCGCTCCA	ACCTGACCGA	CAATCCGCGC	ACCGGACCA
	551	CGTCTGTGCA	CGGTTTCCAC	AATACCGGTA	GTGCTGCTAG	CGAAGAGATA
	601	GGGACGCGAT	TCAAAGCGCG	CCACCGGATC	AGCGCCGAGC	TGGACAGATT
	651	GGCGAATGCC	CGCGAGCTTC	TCCACGCGAC	TGCAGATTTC	GTCAAGATCTC
	701	TCAACGCGAT	CGGCGGAGAA	ATTTCGCGT	CGGCGGAGT	CTCCAGAGGT
30	751	ATAAGCGAAG	CTCCAAACAT	TGCTGTTATG	CACGCTTGG	GTCTGCTTTC
	801	CACCGAAGAA	AGATATGGCG	CCATCAACGA	TGTTCCGAT	ATTGGCGTTC
	851	TCAACAGACTA	TGCGCGACGA	GGCATCCGG	ATTGGGCGAT	CCAAAACCCC
	901	AGTCGGCGAC	AGGCGATCA	AGCGCTCGAC	ATATCTTTA	CGGCGATCT
	951	CCCCCTCAA	GGGATTGGAG	CTGTTCCGGG	AAATACCGG	TTGGCGGGCA
35	1001	TCAACGCGACA	TCCGTCTCAAG	TGCGTCGAGA	TGGGCGAGT	CGCATTCGCG
	1051	AAAGGGGAAT	CGCGCGCTCAG	CGCACAATTT	GGCGATTGGG	CATACGCCAA
	1101	ATCATCGCTTC	CTTACCATTT	CGGCAAAAT	CGTTTCAAC	TTGAGACGAG
	1151	GTATCGGCAA	AGAAACATC	ACCTCTCTCA	CGTTCGGCG	GTCAAACGGT
	1201	AAGATTGTGA	AATCTGCAAA	CAAAACGCA	CGAAGACCA	AAGTCGGGTT
40	1251	TGACGCTTAA	GGGTTTCCGA	ATTTTGA AAA	AGACGTAAAA	ACGATACAGA
	1301	GAATTTAAA	CGGTGTACCA	CAGTGAATC	CTATAGATGA	ATCCGCTGTG
	1351	AATCTTAAT	GTTCCTGCGG	ATCGGCTCAT	CTTCTGGTCTA	TAATCTGAGT
	1401	ATATTAATAC	GTATTAATAC	CGATATATAC	CGATATATAC	TAATCTGAGT
	1451	CTTAAATAC	CTTCTCTTCA	CGACGCTAC	CATTAAGGAC	TAAATATGGA
45	1501	TATTTGGATA	TAATTTGGTA	TGAATGACAT	AAAGTCCAT	CAGAACTAA
	1551	AGGTCAGAAA	TTTGAATGGG	ATGTTCAATT	GAAGTACG	GAAGAGAGCT
	1601	AACTTGGATG	GGGATATAGG	ATGTTGAAGC	ATTTAAATAT	ATCAATTGAT
	1651	GGAAAGATTA	CACCAAAATG	A		

This corresponds to the amino acid sequence <SEQ ID 466>:

50	1	LGSRGSLTSLT	LSLTAVLCFM	HAHASDLAND	STRQVLDRD	HFEFGPKYHL
	51	FSGRGLAEAR	SGSLTGLNTO	SHQLGNLTQ	QAAKNGICY	IVRFSDHGHE
	101	VHSPFFDNHAS	HSDDSDAEGSP	VGGLSFDRHP	WDGHEHHPAD	PDGQFQGGQY
	151	FAPKPGARDIY	SYDLDKVGACN	RLNLTDRHS	TGQRVLDRHD	NTSGMLTGVG
	201	GDGKFKYVY	YVGLDLSGNA	ARAKNLS	YVGLDLSGNA	YVGLDLSGNA
55	251	ISESGNIATV	HQLGLLSTEN	KMARINDLAD	LGGLTQYAAA	ATRDWAVONEV
	301	NAAGQIEAVS	NIFTAVIPVK	GIGAVGRDAP	LMQITAHFVK	RSGMGEGALE
	351	KGKSAVSDNF	ADAAAYAKFS	PPHNSNRFSN	LYDRGKENT	TSTSTVPSNG
	401	KNVLKANLAK	PKTVKPFDEK	GYPNFERKDV	QDTRINTAVP	QVNEIDEPVF
	451	NEPNSGVGSAH	SWSTIATQY	AKFLPQQRIR	YIPPKNYPS	ADGLFRNGSTN
60	501	YLDKFGNWT	KGPSRTKQGE	QEWLDVQKST	GREQLGWASR	DKHINTSTDD
	551	GEIKHS				

Based on this analysis, including the presence of a RGD sequence in the gonococcal protein, typical of adhesins, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 56

- 5 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 467>:

```

1 ATGAATATTC ACACCCCTGCT CTCCAAACAA TGGACGCTGC CGCCATTCCCT
51 GCGGAAACGG CTGCTGCTGT CCTGCTGAT ACTGCTTGCC CCCAATGGCG
101 TGTTTTGGGT TTTGGCACTG CTGACGCCCA CGCCCGCCGC GATTGTCAAT
151 TTGGACTATC TTCCCGCCGC GTGCTGTATC GCCTCGCCTT GCGCTTTGCT
201 CAAATTTGCC GGCCTATTGG CGTTTGGCTT TTTGAAGGGC TTTGAAGGGC
251 TGATGATGGT GATCCAACCT TTCCCTTTTA TGGATCTCAT CGGCGCCATC
301 AACCTCGTCC CTTTCATCCT GACCGCCCCC GCCCCTTATC AGATAATGAC
351 GGGGCTG...

```

This corresponds to the amino acid sequence <SEQ ID 468; ORF48>:

```

1 MNIHTLLSKQ WTLPPFLPKR LLLSLILLIA PNAVFWLAL LTATARPIVN
51 LDYLPALLLI ALPWRVFKIA GVLAFWLAVL FDGLMMVIQL FPFMDLIGAI
101 NLVPFILTAP APYQIMTGL...

```

Further work revealed the complete nucleotide sequence <SEQ ID 469>:

```

1 ATGAATATTC ACACCCCTGCT CTCCAAACAA TGGACGCTGC CGCCATTCCCT
51 GCGGAAACGG CTGCTGCTGT CCTGCTGAT ACTGCTTGCC CCCAATGGCG
101 TGTTTTGGGT TTTGGCACTG CTGACGCCCA CGCCCGCCGC GATTGTCAAT
151 TTGGACTATC TTCCCGCCGC GTGCTGTATC GCCTCGCCTT GCGCTTTGCT
201 CAAATTTGCC GGCCTATTGG CGTTTGGCTT TTTGAAGGGC TTTGAAGGGC
251 TGATGATGGT GATCCAACCT TTCCCTTTTA TGGATCTCAT CGGCGCCATC
301 AACCTCGTCC CTTTCATCCT GACCGCCCCC GCCCCTTATC AGATAATGAC
351 GGGGCTGTTG CTGCTGTATA TGCTGGCGAT GCCCTTTGTG TTGCAGAAAG
401 CCGCGGCCAA AACCCGACTC CGGCACATTC CGCGCTTGCG CGCGCTTTGT
451 CCGCGAGCGG GCTATTTCAC CGGCCATTGG ACTTACTACG ACCGGGGTCG
501 GATGGCCAAT ATCTTCGGCG CAACAACACT CTACTACGCC AAAAGTCAGG
551 CGATGCTCTA CACCGTCAGC CAGAAATCCG ACTTTATTAT CGCCGGCGCTG
601 GTCGATCCCG TCTTCCTCCG CTTGGGCAAT CAACAGCGTG CGGCCACGCA
651 TCTGAACGAG CGGAATCTC AAAAATCCTT CTTTATCGTC CGCGAATCTT
701 GGGGGCTGCC GCGCAATCCC GAACCTCAAA ACGCCACTTT TGCCAAACTG
751 CTGGCGCAAA AAGACCGTTT TTGCGTTTGG GAAAGCGGCA GTTTTCCCTT
801 CATCGGCGCG ACGGTCGAAG CGGAATGCG CGAAGCTGTG GCCTACGCGG
851 GTTTCGCGCG GTTCGCACTG CGCGCGCGCG CGACAGAAAA ATTTGCGCGG
901 TGCCCTCCCA ACCGTTTGAA ACAAGAAGGT TACGCCACCT TTGCGATGCA
951 CGGCGCGGGC AGTTCGCTTT ACGACCGCTT CAGCTGGTAT CGGAGGCGGG
1001 GCTTTCAAGA AATCAAAACC GCGAAAAACC TGATCGGTAA AAAAACCTGC
1051 GCCATTTTCG GCGGCGGTG CGACAGCGAG CTGTTGGCGG AAGTGTGCGG
1101 ATTTTTCAAA AAACACGARA AGGACCTGTT TTACTGGATG ACGCTGACCA
1151 CCACACCGAA CTACCGARA TCCGACAGA TCAACACAG GCTCAAACTG
1201 ACCGAATATG GCCTGCGCGC CGAAACCGAC CTCTGCCCAA ATTTACAGCT
1251 GCACACCCAA TTCTTCGACC AACTGGCGGA TTTGATCCAA CGCCGCGAAA
1301 TGAAGGCGAC GGAAGTCATC ATCGTCGGCG ACCATCCGCG GCCCGTCGGC
1351 AACCTCAATG AACCTTCGCG CTACCTCAAA CAGGGGACAG TCGCTGGCTT
1401 GAACTTCAAA ATCAATAATA

```

This corresponds to the amino acid sequence <SEQ ID 470; ORF48-1>:

```

1 MNIHTLLSKQ WTLPPFLPKR LLLSLILLIA PNAVFWLAL LTATARPIVN
51 LDYLPALLLI ALPWRVFKIA GVLAFWLAVL FDGLMMVIQL FPFMDLIGAI
101 NLVPFILTAP APYQIMTGLL LLYMLAMPEV LQKAAAKTDF RHIAVCAAVV
151 AAAGYFTGHL SYYDRGRMAN IFGANNFYFA KQAMLYTVS QNADFTIAGL
201 VDPVFLPLGN QQRAAATHNE PKSQKILFIV AESGLFANP ELQNAFPAKL
251 LAQKDRFSVW ESGSFFFIGA TVEGEMRELK ATGSLRGFAL RRAPDEKFPAR
301 CLPNRLKQEG YATFAMHGAG SSLYDRFSWY PRAGFQEIKT AENLIGKTKT

```


351 AIFGVCVDSE LFGVSAFFK KHKGLFYWM TLTSHADYPE SDIFNHLKCK
 401 TEYGLDAEDT LGRNFSLHTQ FFDQLADLIQ RPEMKGTEVI IVGDHPPEVG
 451 NLNETFFRYLK QSHVAWLNFK IK*

Computer analysis of this amino acid sequence gave the following results:

5 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF48 shows 94.1% identity over a 119aa overlap with an ORF (ORF48a) from strain A of *N.*

meningitidis:

		10	20	30	40	50	60
orf48.pep		MNIHTLLSKQWTLPPFLPKRLLLSLLTLLAPNAVFVWLALLTATARPVYNLOYLPAALLI					
orf48a		MNIHTLLSKQWTLPPFLPKRLLLSLLTLLAPNAVFVWLALLTATARPVYNLYLPAALLI					
		10	20	30	40	50	60
		70	80	90	100	110	119
orf48.pep		ALPWRVFKIAGVLAFVLAVLFDGLMMVLIQFPFMDLIGAINLVPFITAPAFYQIMTGL					
orf48a		ALPWRVFKIXGLVAXLWLVAVLFDGLMMVLIQFPFMDLIGAINLVPFITAPALYQIMTGLL					
		70	80	90	100	110	120
		130	140	150	160	170	180
orf48a		LLYMLAMPFVLQKAAAKTDFRHHACAAVVVAAGYTGHLXSYDRGRMANIFGANNFYFA					

The complete length ORF48a nucleotide sequence <SEQ ID 471> is:

1	ATGAATATT	ACACCCGTGCT	CTCCAAACAA	TGGACGCTGC	CGCCATTGCT
51	GCGGAAACGG	CTGCTGCTGT	CCTGCTGATG	ACTGCTNNCC	CCCAATGCGG
101	TGTTTGGGT	TTTGCCACTG	CTGACCGCCA	CGCGCGCGCC	GATTGTCAAT
151	TTGGANTACC	TCCCGCGCG	GCTGCTGATC	GCGCTGCGCT	GCGGTNIGCT
201	CAAAATTGNC	GGCGTATTGG	CGTNTGCGCT	GCGGTTTGG	TTGACGCGCG
251	TGATGATGGT	GATCCCACTC	TTCCCTTTTA	TGACATCCAT	CGCGCGCTATC
301	AACTCGTGT	TTTCAATGTT	GACCGCGCG	GCGCTTTATC	AGATAATGAC
351	CGGCTGTGA	CTGCTGTATA	TCGTCGGGAT	GCGGTTTGG	TTGCAGAAAG
401	CCGCGCCCAA	AACGCACTTC	CGACACATTTG	CGCGCTGTGC	CGCGTTTGTG
451	GTGGCAGCCG	GCTATTTTAC	CGGCCATTG	AGTTTANIACG	ACGCGGGGCG
501	GATGGCCAA	ATCTTCGCGG	CAACAACTT	CTATTACGCC	AAAAGTCAGG
551	CGATGCTCTA	CACCGTCAGC	CAGAAATGCG	ACTTTATTAC	CGCGGCGCTG
601	GTGATCCCG	TCTTCTCTCC	CTTGGGCAAT	CAACAGCGTG	CGCGCCACGA
651	TTCTGAACGAG	CGGAAATCTC	AAAAATCCT	CTTTATCGTC	GCGGAATCTT
701	GGGGGCTGCC	GGCCAAATCC	GAACCTCAAA	ACGCGCACTT	TGCCAAACTG
751	CTGGCGCAA	AAGANCGTTT	TTGCGTTTGG	GAAAGCGCGA	GTTTTCCTTT
801	CATCGCGCG	ACGATCGAAG	CGGAAATGCG	CGAAGCTGTG	GCCTACGCGG
851	GTTTTCGCGG	GTTTCGCACTG	CGCGCGCGCG	CGACGAGAAA	ATTTGCGCGG
901	TGCTCTCCCA	ACCGTTTGAA	ACAAGAAGGT	TACGCCACCT	TTGGATGCA
951	CGCGCGCGCG	AGTTGCTTTT	ACGACCGCTT	CAGCTGGTAT	CGGAGGCGCG
1001	GCTTTCAAGA	AATCAAAACC	GCGGAAACCC	TGATCGGATA	AAAACCTGAT
1051	GCACTTTTC	GCGCGCTGTG	CGACAGCGAG	CTGTGTCGCG	AAGTGTGCTG
1101	NTTTTTCAAA	AAACACGACA	AGGACTGTTT	TTACTGGATG	ACGCTGACCA
1151	CCGACGCGGA	CTATCCCGAA	TCNCACTTTT	TCAACACAG	GCTCAATGTC
1201	ACGGAATATG	GCCTGCCCGC	CGAAACCGAC	NPTCGCCGCA	ATTTGACGCT
1251	GCACACCCAA	TTCTTGGACC	AACTGGCGGA	TTTGATCCAA	CGCGCGCGAA
1301	TGAAGGCAC	GGAAGTCATC	ATCGTGGCGG	ACCATCCGCG	GCGCGTCGCG
1351	AACTCAATG	AAACCTTCGG	CTACCTCAAA	CAGGGGACAG	TGCTGTCGCT
1401	GAACCTCAAA	ATCAAAATAA			

This encodes a protein having amino acid sequence <SEQ ID 472>:

1	MNIHTLLSKQ	WTLPPFLPKR	LLLSLLILX	PNAVFVWLAL	LTATARPVYN
51	LXYLPAAALLI	ALPWRVFKIX	GVLAAXWLAVL	FDGLMMVLIQ	FPFMDLIGAI
101	NLVPFFIXTAP	ALYQIMTGLL	LLYMLAMPFV	LQKAAAKTDF	RHHACAAV
151	VAAGYFTGHL	SKYDRGRMAN	IFGANNFYFA	KSQAMLYTVS	QNADFITAGL
201	VDVPFLPLGN	QQRATHLNE	PKSQKILTV	AEWGLFPANF	ELQWATFAKL
251	LAKKMFVYM	ESGSPFLGA	TTGEMRBLK	ATGGLRGLR	RRAPDEKAT
301	CLFNRLKQEG	YATFAMHAG	SSLYDRFSWY	PRAGFQKKT	ABNLGKTKT
351	AIFGVCVDSE	LFGVSAFFK	KHKGLFYWM	TLTSHADYPE	SDIFNHLKCK

401 TEYGLPAETD KCRNFSLHTQ FFDQLADLIQ RFEMKGTVEI IVGDHPPVPV
 451 NLNETFRYLK QGHVXWLNFK IK*

ORF48a and ORF48-1 show 96.8% identity in 472 aa overlap:

		10	20	30	40	50	60
5	orf48a.pep	MNIHTLLSKQWTLPPFLPKRLLSLLILLAPNAVFWVLALLTATARPIVNLXYLPAALLI					
	orf48-1	MNIHTLLSKQWTLPPFLPKRLLSLLILLAPNAVFWVLALLTATARPIVNLXYLPAALLI					
		10	20	30	40	50	60
10	orf48a.pep	ALPWRXVKIXGVLAFLAVLFDGLMMVIQLFFMDLIGAINLVFILTAPAPYQIMTG					
	orf48-1	ALPWRXVKIXGVLAFLAVLFDGLMMVIQLFFMDLIGAINLVFILTAPAPYQIMTG					
		70	80	90	100	110	120
15	orf48a.pep	LLYMLAMPFVLQKAAAKTDFRHIAACAIVVAAGYFTGHLSDYDRGRMANIFGANNFYA					
	orf48-1	LLYMLAMPFVLQKAAAKTDFRHIAACAIVVAAGYFTGHLSDYDRGRMANIFGANNFYA					
		130	140	150	160	170	180
20	orf48a.pep	KSQAMLYTYSQNADFITAGLVDFVFLPLGNQQRAATHLNEPKSQKILFIVAESWGLFANP					
	orf48-1	KSQAMLYTYSQNADFITAGLVDFVFLPLGNQQRAATHLNEPKSQKILFIVAESWGLFANP					
		190	200	210	220	230	240
25	orf48a.pep	ELQNAATFAKLLAQKXRFVSWESGSGFFIGATIEGEMRELCAYGGLRGFALRRAPDEKFA					
	orf48-1	ELQNAATFAKLLAQKXRFVSWESGSGFFIGATIEGEMRELCAYGGLRGFALRRAPDEKFA					
		250	260	270	280	290	300
30	orf48a.pep	CLPNRLKQEGYATFAMHGAGSSLYDRFSWYPRAGFQEKTAENLIGKKTCAIFGGVCDSE					
	orf48-1	CLPNRLKQEGYATFAMHGAGSSLYDRFSWYPRAGFQEKTAENLIGKKTCAIFGGVCDSE					
		310	320	330	340	350	360
35	orf48a.pep	LFGEVSAXFKKHKDGLFYWMTLTSHADYPESDIFNHLKCTEYGLPAETDLCRNFSLHTQ					
	orf48-1	LFGEVSAXFKKHKDGLFYWMTLTSHADYPESDIFNHLKCTEYGLPAETDLCRNFSLHTQ					
		370	380	390	400	410	420
40	orf48a.pep	FFDQLADLIQRPEMKGTEVIIIVGDHPPFVGNLNETFRYLKQGHVXWLNFKIKX					
	orf48-1	FFDQLADLIQRPEMKGTEVIIIVGDHPPFVGNLNETFRYLKQGHVXWLNFKIKX					
		430	440	450	460	470	
45	orf48a.pep						
	orf48-1						
50	orf48a.pep						
	orf48-1						

Homology with a predicted ORF from *N.gonorrhoeae*

ORF48 shows 97.5% identity over a 119aa overlap with a predicted ORF (ORF48ng) from *N.gonorrhoeae*:

55	orf48.pep	MNIHTLLSKQWTLPPFLPKRLLSLLILLAPNAVFWVLALLTATARPIVNLXYLPAALLI	60
	orf48ng	MNIHTLLSKQWTLPPFLPKRLLSLLILLAPNAVFWVLALLTATARPIVNLXYLPAALLI	60
60	orf48.pep	ALPWRXVKIXGVLAFLAVLFDGLMMVIQLFFMDLIGAINLVFILTAPAPYQIMTG	119
	orf48ng	ALPWRXVKIXGVLAFLAVLFDGLMMVIQLFFMDLIGAINLVFILTAPAPYQIMTG	120

The ORF48ng nucleotide sequence <SEQ ID 473> was predicted to encode a protein having amino acid sequence <SEQ ID 474>:

```

1  MNHALLSEQ WTLPPFLPKR LLLSLILLIA PNAVFVLIAT LTATARPIVN
51 LDYLPALLI ALPWRFKVIA GVLAFWPAVL FDGLMMVIQL FPFMDLIGAI
101 NLVPFILTAP APYQIMTGLL LLYMLAMPFV LQKAAVKTFD RHIAVCAAVV
151 AAARYFTGPF ELLRTGGRWQ YVQHRRLLS GSRAFRRRQ KADVLRLLGN
201 PYASMGNGG..

```

Further work identified the complete gonococcal DNA sequence <SEQ ID 475>:

```

1  ATGAATATTC ACGCCCTGCT CTCGGAACAA TGGACGCTGC CGCCATTCTCT
51 GCGCAAAACGG CTGCTGCTGT CCTGCTGAT ACTGCTGCGC CCCAATGCGG
101 TGTTTTGGGT TTTGGCACTG CTGACCGCCA CCGCCCGCCC GATTTCATAT
151 TTGGACTACC TTCCGCGCGG GCTGCTGATC GCGCTGCTGT GCGCTTCTGT
201 CAAAATTGCC GCGCTATTGG CTTTGTGGCC GCGGCTTTTG TTTGACGGGC
251 TGTATGAGTG GATGCAACTG TTCCCTTTTA TGACACTCAT CGCGCGCATC
301 AACCTGCTGC CTTTCATCCT GACCGCCCCC CCGCCTTATC AGATATATGAC
351 CGGGCTGTGT CTGCTGTATA TGCTGGCGAT GCGCTTTGTG TTGCAAAAAG
401 CGCGCGTCAA AACCGACTTC CGACACATTG CCGTCTGTGC CGCGCTTGTG
451 CGCGCAGCGG GCTATTTTAC CGGCCATTGT AGTTACTACG ACCGGGGGCG
501 GATGGCCAAT ATCTTCGGCG CAACAACCTT CTATTACGCG aAAAGTCAGG
551 CGATGCTCTA CACGTCAGC CAGAATGCCG ACTTTATTAC CGCCGgctGT
601 GTCGACCCCG TCTTCTCCCC CTTGGGCAAT CAGCAGCGTG CGCGCACGCG
651 GCTGAGTGAG CCGAATCTC AAAAATCCT CTTTATCGTC GCCGAATCTT
701 GGGGGCTGCC GGGCAATCCC GAGCTTCAAA ACGCACTTT TGCCAAACTG
751 CTGGCGCAAA AAGACCGTTT TTCGTTTGG GAAAGCGGCA GTTTTCCCTT
801 CATCGGCGCG ACGGTCGAAG GCGAAATGCG CGAATTGTGC GCCTACGCGG
851 GTTTCGCGCG GTTCGCACGT CGCCGCGCGC CGACGAAAA ATTTGCCCCG
901 TGCCCTCCCA ACCGTTTGAA ACAAGAAGGT TACGCCACTT TTGCGATGCA
951 CGCGCGGGGT AGTTCGCTAT ACGACCCCTT CAGCTGATAT CCGAGGCGCG
1001 GCTTTCARAA ATCAAAACG CCGCAAAACC TGATCGGTAA AAAAACTTGC
1051 GCAATTTTCG GCGCTGTG GACACGAGG CTGTTGCGCG AAGTGTCCGC
1101 ATTTTTCAAA AAACACGACA AGGGACTGTT TTACTGGATG ACGCTGACCA
1151 GGCACGCCGA CTATCCCGAA TCCGACATTT TCAACACAG GCTCAATGTC
1201 ACCGAATACG GCCTGCCCGC CGAAACCGAC CTCTGCCGCA ATTTACAGCT
1251 GCACACCCAA TtcttcgACC AACTGGCGGA TTTGATCCGA CGCCCGCAAA
1301 TGAAGGCGAC GGAAGTCATC ATCGTCCGGC ACCATCCGCG GCCCGTCCGC
1351 AACCTCAATG AACCTTCCG CTACCTCAAA CAGGGACAGC TCGCGTGGCT
1401 GCACCTTCAA ATCAATAA

```

This encodes a protein having amino acid sequence <SEQ ID 476; ORF48ng-1>:

```

1  MNHALLSEQ WTLPPFLPKR LLLSLILLIA PNAVFVLIAT LTATARPIVN
401 LDYLPALLI ALPWRFKVIA GVLAFWPAVL FDGLMMVIQL FPFMDLIGAI
101 NLVPFILTAP APYQIMTGLL LLYMLAMPFV LQKAAVKTFD RHIAVCAAVV
151 AAAGYFTGHL SYYDRGRMAN IFGANNFYVA KSQAMLYTVS QNADFITAGL
201 VDPVFLPLGN QQRATRLISE FKSKLLEIV AEWGLPQNF ELQVATFRL
251 LAQKDRFVIV EGGTFPFGA IVGEMRELK AYGLGAFAL RAEDEKFAK
301 CLPNLRKQEG YATFAMHGAG SSLYDRFSWY PRAGFKIKT AENLICKKTC
351 AIFGVGCDSE LFGVSAFFK KHKDKLFYWM TLTSHADYFE SDIFNRLKLC
401 TEYGLPAETD LCRNFSLHTQ FFDQLADLIR REPMKGTVEI IVGDHPFPVG
451 NLNETERYLK QHVAWLHEK IK*

```

ORF48ng-1 and ORF48-1 show 97.9% identity in 472 aa overlap:

50		10	20	30	40	50	60
	orf48-1.pep	MNHTLLSKQW	TLPPFLPKR	LLLSLL	LLAPNAVFVLI	ALLTATARPIVN	LDYLPALLI
	orf48ng-1	MNHALLSEQW	TLPPFLPKR	LLLSLL	LLAPNAVFVLI	ALLTATARPIVN	LDYLPALLI
55		70	80	90	100	110	120
	orf48-1.pep	ALPWRFKVIAGV	LAFLAVLFDGLMMVIQL	FPFMDLIGAIN	LVLPFILTAP	APYQIMTGLL	
	orf48ng-1	ALPWRFKVIAGV	LAFLAVLFDGLMMVIQL	FPFMDLIGAIN	LVLPFILTAP	APYQIMTGLL	
60		70	80	90	100	110	120

		130	140	150	160	170	180
	orf48-1.pep	LLYMLAMPFVLQKAAAKTDFRHI	AVCAAVVAAAGYFTGHL	SYDDGRGMANIFGANNFYFA			
5	orf48ng-1	LLYMLAMPFVLQKAAVKTDFRHI	AVCAAVVAAAGYFTGHL	SYDDGRGMANIFGANNFYFA			
		130	140	150	160	170	180
	orf48-1.pep	KSQAMLYTVSQNADFITAGLVDP	VFPLPLGNQQR	AATHLNEPKSQKILFVAESWGLP	PANP		
10	orf48ng-1	KSQAMLYTVSQNADFITAGLVDP	VFPLPLGNQQR	AATHLNEPKSQKILFVAESWGLP	GNP		
		190	200	210	220	230	240
	orf48-1.pep	ELQNATFAKLLAQKDRFSVWESGS	FFFIGATVEGEMREL	CAYGGLRGFALRRAPDEK	FFAR		
15	orf48ng-1	ELQNATFAKLLAQKDRFSVWESGS	FFFIGATVEGEMREL	CAYGGLRGFALRRAPDEK	FFAR		
		250	260	270	280	290	300
	orf48-1.pep	CLPNRLKQEGYATFAMHGAGSS	SLYDRFSYPFRAGFQ	EIKTAENLIGKKTCAIFGGV	CDSE		
20	orf48ng-1	CLPNRLKQEGYATFAMHGAGSS	SLYDRFSYPFRAGFQ	EIKTAENLIGKKTCAIFGGV	CDSE		
		310	320	330	340	350	360
	orf48-1.pep	LFGEVSAPFFKKHDKGLFYWM	TLTSHADYPESDIFN	HRLKCTEYGLPAETDL	CRNFS	SLHTQ	
25	orf48ng-1	LFGEVSAPFFKKHDKGLFYWM	TLTSHADYPESDIFN	HRLKCTEYGLPAETDL	CRNFS	SLHTQ	
		370	380	390	400	410	420
	orf48-1.pep	FFDQLADLIRPEMKGT	EVIIIVGDHPPPVGN	LNETFRLYKQGHVAVL	NFKIKX		
30	orf48ng-1	FFDQLADLIRPEMKGT	EVIIIVGDHPPPVGN	LNETFRLYKQGHVAVL	NFKIKX		
		430	440	450	460	470	
	orf48-1.pep	FFDQLADLIRPEMKGT	EVIIIVGDHPPPVGN	LNETFRLYKQGHVAVL	NFKIKX		
35	orf48ng-1	FFDQLADLIRPEMKGT	EVIIIVGDHPPPVGN	LNETFRLYKQGHVAVL	NFKIKX		

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and two putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 57

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 477>:

```

1  ..GTGACGGGAC GTTACCGCGC TTTGGATCGC GTTTCACAAA TCATCATCGT
51  TACTTTGAGT ATCGCCACGC TTGCGCGCGC CGGCATCGCT ATGTCGCGGG
101 GTATGCAGAT GCAGTCCGAT TTTATCGAGC CGACACCGTG GACGCTTGGC
151 GGTTTGGGCT TCCTGATCGC GCTGATGGGC TGGATGCCCG CGCCGATTA
201 AATTTCCGCC ATCAATTCTT TGTGGGTAA CCGAAAACAA CGCATCAATG
251 CTTCGGAATA CCGCGACGGG ATTTTGTGA TCAACGTCGG TTATATCGCC
301 AGTGCGGTTT TGGCTTTGGT TTTCTTGA CTTGGCGC G TAGCGCGAA
351 CGGCAACGGC GA.ACAGTGC AGATGGCGGG CGGCAAAAT AACGGCGCAAT
401 TGATCAATAT GTACGCC..

```

This corresponds to the amino acid sequence <SEQ ID 478; ORF53>:

```

1  ..VSGRYRALDR VSKIIIVTSL IATLAAAGIA MSRGMQMQSD FIEPTPTWLA
51  GLGLFIALMG WMPAPIEISA INSLWTEKQ RINPSEYRDG IFEFNVGYIA
101 SAVLALVFLA LGXVAFNSNG XTVMAGGKY NGQLINMTA..

```

Further work revealed the complete nucleotide sequence <SEQ ID 479>:

```

1  ATGTCGGAAC AACATATTTT GACTTGGAAA AGTAAATCA ACGCATTGGG

```

5 251 TCCGGGGATC ATGATGGCTT CGGCGCGCGT CGGCGGTTGC CACCTGATTG
 101 CCTCGAGCGA GCGGCGCGCG CTTTACGGCT GGCAGATCGC GCTCATCATC
 151 ATCTTGACCA ACCTCTTCAA ATACCCGTTT TTCCGCTTCA GCGCGGATTA
 201 CACGCTGGAC ACGGGCAAGA GCGTATTGTA AGGTATATGCC GAGAAAAGCC
 251 GCGTTTATTT GTGGGTATTC CTGATTTTGT GCATCTCTC CGCCACGATT
 301 AACCGGGGCG CGGTGCGCAT TGTAACCGCC GCCATCGTCA AAATGCGCAT
 351 TCCCTCGCTG ATGTTTGTAT GCGGCACGGT TGCCTGCTTG ATTATGGCAT
 401 CTTGCTGTAT TATTTTGTGT AGCGGACGTT ACCGCGCTTT GGATCGCGTT
 451 TCCAAATATCA TCATCGTTAC TTTGAGTATC GCCACGCTTG CCGCGCGCGG
 10 501 CATCGCTATG TCGGCGCGTA TGCAGATGCA GTCCGATTTT ATCGAGCGCA
 551 CACCGTGGAC GCTTGCGGTT TTGGGCTTCC TGATCGCGCT GATGGCTTGG
 601 ATGCGCGCGC CGATTGAAAT TTCCGCCATC AATTCTTTGT GGGTAACCGA
 651 AAAACACAGC ATCAATCCTT CGGAATACCG CGACGGGATT TTTGATTCCA
 701 ACGTCGGTTA TATCGCCAGT GCGGTTTGTG CTTTGGTTTT CTTTGCACGT
 15 751 GCGCGCTTTG TGCAATACGG CAACGGCGAA CAGTGCAGA TGGCGCGCGG
 801 CAAATATATC GGGCAATGTA TCAATATGTA CGCGCTTACC ATCGGCGCGT
 851 GGTGCGGCCC GCTGGTGGCG TTTATCGCGT TTGCTGTAT GTACGSCACG
 901 ACGATTACCG TCGTGGACGG CTATGCGCGT CGCATTTGCC AACCGTGGC
 351 CCGTCTGCGC GGAAGAACA AACGGCGGTA CGCGGAATTC TTTGCTGGA
 20 1001 ATATTTGGGT GCGGCGCAGC GGTTCGCGG TGATTTCTG GTTGACGGC
 1051 GTATTGGCGA ATCTGCTCAA ATTTGCGATG ATTGCGCGTT TTTGTCGCGC
 1101 CCTGTGTGTT GCTTGGCTGA ATTACCGTTT GGTAAAGGT GATGAAGAAC
 1151 ACAARCTCAC ATCAGGTATG AATGCCCTTG CATTTGCAGG CTTGATTAT
 25 1201 CTGACCGGTT TTACCGTTT GTTCTTATTG AATTGGCGG GAATGTTCAA
 1251 ATGA

This corresponds to the amino acid sequence <SEQ ID 480; ORF53-1>:

1 MSEQHISTWK SKINALGPFI MMASAAVAGS HLIASTQAGA LYGWQIALII
 51 ILTNLFKYPF FRFSAHYTLT TGKSLIEGYA EKSRYVLWVF LILCILSATI
 30 101 NAGAVAIVTA AIVKMAIPSL MFDAGTVAAL IMASCLIIIV SGRYRALDRV
 151 SKIIIVTLSI ATLAAAGTAM SRGMQMSDF IEPTFWTLAG LGFLIALMGW
 201 MPAPIEISAI NSLWVTEKQR INPSEYRDI FDFNVGYIAS AVLALVFLAL
 251 GAFVQYNGNE AVQMAGGKYI GQLINMYAVT IGGWSRFLVA FIAFACMYGT
 301 TITVVDGYAR AIAEPVRLLR GKDKTGNAEF FAWNIVWAGS GLAVIFWFDG
 351 VMANLKFAM IAAFVSAFVF AWMLYALVKG DEKHILTSGM NALALAGLIV
 35 401 LTGFTVLFLL NLAGMFK*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF53 shows 93.5% identity over a 139aa overlap with an ORF (ORF53a) from strain A of *N. meningitidis*:

40 orf53.pep VSGRYRALDRVSKIIIVTLSIATLAAAGIA
 orf53a AAVIKMAIPSLMFDAGTVAALIMASCLIIIVSGRYRALDRVSKIIIVTLSIATLAAAGIA
 45 110 120 130 140 150 160
 orf53.pep MSRGMQMSDFIEPTFWTLAGLFLIALMGWMPAPIEISAINSLWVTEKQRINPSEYRDG
 orf53a MSRGMQMSDFIEPTFWTLAGLFLIALMGWMPAPIEISAINSLWVTEKQRINPSEYRDG
 50 170 180 190 200 210 220
 orf53.pep IFEFNVGYIASAVLALVFLALGXVAPNGNXTVMAGGKYNGQLINMYA
 orf53a IFEFNVGYIASAVLALVFLALGAFVQYNGNEAVQMAGGKYIGQLINMYAVTIGGWSRPLV
 55 230 240 250 260 270 280
 orf53a AFIACFACMYGTITITVVDGYARAIEPVRLLRGKDKTGNAEFFAWNIVWAGSGLAVIFWFD
 290 300 310 320 330 340

The complete length ORF53a nucleotide sequence <SEQ ID 481> is:

1 ATGTCGGAAC AACATATTTT GACTTGGAAA AGTAAATCA ACGCATTGGG
 51 ACGGGGGATT ATCATGGGCTT CGGCGGGCGT CGGCGGGTTC CACCTGATTG
 101 CTGCGAGCA GCGCGGGCGG CTTTACGGCT GCGAGCATCG CACTCATCAT
 151 ATCTGTGACA ACCTCTTCAA ATACCGGTTT TTCGCGTTCA GCGCGGATTA
 5 201 CACGCTGGAC ACGGGCAAGA GCCTGATTGA AGGTTATGCC GAGAAAAGCC
 251 GCGTTTATTT GTGGGTATTC CTGATTTTGT GCATCCTCTC CGCCACGATT
 301 AACGCGGGCG CGGTGCGCAT TGTAACGCCG GCCATCGTCA AATGCGCGAT
 351 TCCCTCGCTG ATGTTTGATG CCGGCACGGT TGCCGCGCTT ATTATGGCAT
 401 CCGTCCCTGAT TATTTTGGTG AGCGGACGTT ACGCGCGTTT GGATCGGCTT
 451 TCCAAATATCA TCATCGTTAC TTTGAGTATC GCCACGCTTG CGCGCGCGCG
 501 CATCGCTATG TCGCGCGGTA TGCAGATGCA CTCGAGTTT ATCGAGCGGA
 551 CACCGTGGAC GCTTGCCGGT TTGGGCTTCC TGATCGGCGT GATGGGCTGG
 601 ATGCCCCGCG CGATTGAAAT TTCGCCCATC AATTTCTTGT GGGTAACCGA
 651 AAAACACACG ATCAATCCTT CGGAATACCG CGACGGGATT TTTGATTATT
 701 ACGTCGGTTA TATCGCCAGT GCGGTTTTGG CTTTGGTTTT CCTTGCACTG
 751 GGGCGGTTTT TGCAATACGG CAACGGCGAA GCACTGCAAG TGGCGGGCGG
 801 CAAATATATC GGGCAATTGA TCAATATGTA CGCCGTATAC ATCGCGCGCT
 851 GGTGCGCCCC CTTGTGGCGG TTTATCGGCT TTGCTGTATG GTACCGCACG
 901 ACGATTACCG TTGTGGACGG CTATGCGCGT GCCATTGCC AACCGGTGGG
 951 CTGCTGCGCC GGAAGAAGACA AACCGGCCAA CGCCGAATTC TTTGCTGGA
 1001 ATATTGGGCT GCGCGGCAGC GGTTTGCGGG TGATTTTCTG GTTTGACCGC
 1051 GTAATGCGCA ATCTGCTCAA ATTTGCGATG ATTGCGGCTT TTGTGTCGCG
 1101 CCTGTGTTT GCCTGGCTGA ATTACGGTTT GGTCAAAGGT GATGAAAAAC
 1151 ACAAACCTAC ATCAGGTATG AATGCGCTTG CATTGCGAGG CTTGATTTAT
 1201 CTGACCGGTT TTACCGTTTT GTTCTTATTG AATTGGCGGG GAATGTTCAA
 1251 ATGA

This encodes a protein having amino acid sequence <SEQ ID 482>:

1 MSEQHISTWK SKINALGPGI MMASAAVGGG HLIASTQAGA LYGWQIALII
 51 ILTNLFKYFP FRFSAHYTLD TGKSLIEGYA EKSRYLVWF LILCILSATI
 101 NAGAVAIVTA AIVKMAIPSL MFDAGTVAAL IMASCLIIIV SGRYRALDRV
 151 SKIIIVTLSI ATLAAGTAM SRGMQMSDF IEPTFWTLAG LGFLIALMGW
 201 MPAPIEISAT NSLWVTEKQR INFSEYRDGI FDFNVGYIAS AVLALVFLAL
 251 GAFVQYNGE AVQMAGGKYI QGLINMYAVT IGGWSRPLVAF AFACMYGT
 301 TIVFLALGAF AAFVSRPVL RKRDTGNAEF FARNIVVAGS GLAVIWFDFG
 351 VMANLKFAM IAAFVSARVF AWINYRLVKG DEKRHTTSCM NALALAGLIY
 401 LTGFTVLELL NLAGMPK*

ORF 53a shows 100.0% identity in 417 aa overlap with ORF53-1:

	10	20	30	40	50	60
orf53a.pep	MSEQHISTWKS	SKINALGPGIM	MASAAVGGSH	LIAS	TQAGALY	GWQIALII
orf53-1	MSEQHISTWKS	SKINALGPGIM	MASAAVGGSH	LIAS	TQAGALY	GWQIALII
	10	20	30	40	50	60
orf53a.pep	FRFSAHYTLD	TGKSLIEGYA	EKSRYLVWF	LILCILSATI	NAGAVAIVTA	AIVKMAIPSL
orf53-1	FRFSAHYTLD	TGKSLIEGYA	EKSRYLVWF	LILCILSATI	NAGAVAIVTA	AIVKMAIPSL
	70	80	90	100	110	120
orf53a.pep	MFDAGTVAAL	IMASCLIIIV	SGRYRALDRV	SKIIIVTLSI	ATLAAGTAM	SRGMQMSDF
orf53-1	MFDAGTVAAL	IMASCLIIIV	SGRYRALDRV	SKIIIVTLSI	ATLAAGTAM	SRGMQMSDF
	130	140	150	160	170	180
orf53a.pep	IEPTFWTLAG	LGFLIALMGW	MPAPIEISAT	NSLWVTEKQR	INFSEYRDGI	FDFNVGYIAS
orf53-1	IEPTFWTLAG	LGFLIALMGW	MPAPIEISAT	NSLWVTEKQR	INFSEYRDGI	FDFNVGYIAS
	190	200	210	220	230	240
orf53a.pep	AVLALVFLAL	GAFVQYNGE	AVQMAGGKYI	QGLINMYAVT	IGGWSRPLVAF	AFACMYGT
orf53-1	AVLALVFLAL	GAFVQYNGE	AVQMAGGKYI	QGLINMYAVT	IGGWSRPLVAF	AFACMYGT
	250	260	270	280	290	300

		310	320	330	340	350	360
5	orf53a.pep	TTTVVDGYARAIAEPVRLLRGKDKTGNAEFFAWNIVWAGSGLAVIWFWDGVMANLLKFAM					
	orf53-1	TTTVVDGYARAIAEPVRLLRGKDKTGNAEFFAWNIVWAGSGLAVIWFWDGVMANLLKFAM					
		310	320	330	340	350	360
10	orf53a.pep	IAAFVSAPVFAWLNYRLVKGDEKHLTSGMNALALAGLIYLTGTFTVLFLNLAGMFKX					
	orf53-1	IAAFVSAPVFAWLNYRLVKGDEKHLTSGMNALALAGLIYLTGTFTVLFLNLAGMFKX					
		370	380	390	400	410	

Homology with a predicted ORF from *N.gonorrhoeae*

- 15 ORF53 shows 92.1% identity over a 139aa overlap with a predicted ORF (ORF53ng) from *N.gonorrhoeae*:

	orf53.pep	VSGRYRALDRVSKIIIVTSLIATLAAAGIA	30
20	orf53ng	AAIVKMAIPSLMFDAGTVAALIMASCLIIIVSGRYRALDRVSKIIIVTSLIATLAAAGIA	91
	orf53.pep	MSRGMQMQSDFIEPTFWTLAGLGLFIALMGWMPAPIEISAINSLVWTEQRINPSEYRDG	90
	orf53ng	MSRGMQMQSDFIEPTFWTLAGLGLFIALMGWMPAPIEISAINSLVWTEQRINPSEYRDG	151
25	orf53.pep	IFEFNVGYIASAVLALVFLALGXVAPNGNGXTVMAGGKYNGQLINMYA	139
	orf53ng	IFDENVGYIASAVLALVFLALGAFVQYNGEAVQMGGKYIGQLINMYAVNTIGGSRPLV	211

An ORF53ng nucleotide sequence <SEQ ID 483> was predicted to encode a protein having amino acid sequence <SEQ ID 484>:

	1	MEKKSCVYLW VFLILCTASA TINAGAVAIV TAAIVKMAIP SLMFDAGTVA
	51	ALIMASCLII LVSGRYRALD RVSKIIIVTL SIATLAAAGI AMSRGMQMOP
	101	DFIEPTFWTL AGLGLFIALM GWMMPAPIEIS AINSLVWTE QRINPSEYRD
	151	GIFDENVGYI ASAVIALVFL ALGAFVQYNG GEAVQMGGK YIGQLINMYA
35	201	VTIGGGSRPL VAFIAFACMY GAASTVVDGY ARAIAEPVRL LRKDKTARP
	251	IVLLEKLGR HRFGRDFLV*

Further analysis revealed further partial DNA gonococcal sequence <SEQ ID 485>:

	1	..aagaAAAGCT GCGTTTATTT GTGGGTTTTT TTGATTTTGT GTATCGCCTC
40	51	CGCCACGATT AACGCGGGCG GCGTCGCCAT TGTACCGCC GCCATCGTCA
	101	AAATGGCGAT TCCTCGCTG ATGTTTGATG CCGGCACGCT TGCCGCTTG
	151	ATTATGGCAT CTGCTCTGAT TATTTTGGTG AGCGGACGTT ACCGCGCTTT
	201	GGATCGTGT TCAGAAATCA TCATTGTTAC TTTGAGCATC GCACGCTTG
	251	CCGCGCGCGG CATCGCTATG TCGCGCGGTA TGCAGATGCA GCCCGATTTT
45	301	ATCAGAGCGA CACGCTGGAC GCTTGCCGGT TTGSECTTCC TGATCGCGCT
	351	GATGGGCTGG ATGCGCGCGC CGATCGAAT TTGCGCCAPC AATCTTGTT
	401	GGGTAAACGA AATACACGAG ATCAATCCTT CTGAAATACG CGACGGGATT
	451	TTGCAATTCA ACGTCGGTTA TATCGcaag GCGTTTGTG CTTTGGTTTT
	501	CCTTGCACTG GCGCGGTTT TGCAATACGG CAACGCGCAA CGAGTGCAGA
	551	TGGCGGGCGG CAAATATATC GGGCAATTGA TTAATATGTA TGCCGTAAAC
50	601	ATCGGGCGCT GGTCTGCTCC GCTGGTGGCG TTTATCGCGT TTGCTCTGAT
	651	GTACGGCAGC ACGATTACCG TTGTGACGCG TTATGCGCGT GCCATTGCGG
	701	AACCGCTGCG CTGCTGCGCG GCGAGGGATA AAACCGGCNA CGCCAGATTG
	751	TTTgctTGGA ATATTGGGT GCGGGGCGAG GGTTTGGCGG TGATTTTCTG
	801	GTTTGACggc gcaatGGCgG AACTgcTCAA ATTTGCGGAT ATtgccgcCT
55	851	TTGTGTCCGC CCTGTGTTTC GCTTGCTCAA ACTACGCGCT CGTCAAGGG
	901	GACAAACGCC ACAGGCTTAC CGCCGGTATG AACGCGCTTG CCATTGTGCG
	951	CCTGCTCTAC CTGGCGGGGT TTGCGGTTTT GTTCTGTTG AACCTACCG
	1001	GACTTTTGGC ATAG

This corresponds to the amino acid sequence <SEQ ID 486; ORF53ng-1>:

```

1  ..KKSCVYLVWF LILCIASATI NAGAVAIVTA AIVKMAIPSL MFDAGTVAAL
51 IMASCLIIIV SGRYRALDRV SKIIIVTISI ATLAAGIAM SRGMQMOPDF
101 IEPTFWTLAG LGFLIALMGW MPAPITISAI NSLWVTEKQR INPSEYRGI
151 FDFNVGYIAS AVIALVFLAL GAFVQYNGE AVQMAGGKYI GQLINMYAVT
201 IGWWSRPLVA FIAFACMYGT TITVVDGYAR AIAEPVRLLR GRDKTGNAEL
251 FAWNIVVAGS GLAVIFWFDG AMAELKFAM IAAVFSAPVF AWINYRLVKG
301 DKRHRLTAGM NALAIVGLLY LAGFAVLFL NLTGLLA*

```

ORF53ng-1 and ORF53-1 show 94.0% identity in 336 aa overlap:

```

10      orf53-1.pep  60      70      80      90      100      110
      orf53ng-1    ILTNLFKYPPFRFSAHYTLDTGKSLIEGYAEKSRVYLWVFLILCILSATINAGAVAIVTA
                        :||| ||||| ||||| ||||| ||||| |||||
                        KKSCVYLVWVFLILCIASATINAGAVAIVTA
                        10      20      30

15      orf53-1.pep  120      130      140      150      160      170
      orf53ng-1    AIVKMAIPSLMFDAGTVAALIMASCLIIIVSGRYRALDRVSKIIIVTSLIATLAAAGIAM
                        ||||| ||||| ||||| ||||| ||||| |||||
                        AIVKMAIPSLMFDAGTVAALIMASCLIIIVSGRYRALDRVSKIIIVTSLIATLAAAGIAM
                        40      50      60      70      80      90

20      orf53-1.pep  180      190      200      210      220      230
      orf53ng-1    SRGMQMOPDFIEPTFWTLAGLGLFLIALMGWMPAPIEISAINSLMWTEKQRINPSEYRGI
                        ||||| ||||| ||||| ||||| ||||| |||||
                        SRGMQMOPDFIEPTFWTLAGLGLFLIALMGWMPAPIEISAINSLMWTEKQRINPSEYRGI
                        100     110     120     130     140     150

25      orf53-1.pep  240      250      260      270      280      290
      orf53ng-1    FDFNVGYIASAVLALVFLALGAFVQYNGEAVQMAGGKYIGQLINMYAVTIGGWSRPLVA
                        ||||| ||||| ||||| ||||| ||||| |||||
                        FDFNVGYIASAVLALVFLALGAFVQYNGEAVQMAGGKYIGQLINMYAVTIGGWSRPLVA
                        160     170     180     190     200     210

30      orf53-1.pep  300      310      320      330      340      350
      orf53ng-1    FIAFACMYGTTITVVDGYARAIAEVPVRLLRGRDKTGNAELFAWNIVVAGSGLAVIFWFDG
                        ||||| ||||| ||||| ||||| ||||| |||||
                        FIAFACMYGTTITVVDGYARAIAEVPVRLLRGRDKTGNAELFAWNIVVAGSGLAVIFWFDG
                        220     230     240     250     260     270

35      orf53-1.pep  360      370      380      390      400      410
      orf53ng-1    VMANLLKFAMIAAEVFSAPVFAMWINYRLVKGDEKHKLTSGMNALAGLIYLTGTVLFL
                        :||: ||||| ||||| ||||| ||||| |||||
                        AMAELKFAMIAAEVFSAPVFAMWINYRLVKGDKRHRLTAGMNALAVGLIYLAGFAVLFL
                        280     290     300     310     320     330

40      orf53-1.pep  280      290      300      310      320      330
      orf53ng-1    NLAGMFEKX
                        ||:|:|
                        NLTGLLAX

45      orf53-1.pep  NLAGMFEKX
                        ||:|:|
                        NLTGLLAX

```

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 58

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 487>:

```

1  ..TTGCGGGAAA CGGCATATGT TTTGGATAGT TTTGATCGTT ATTTTGTGTG
51 TGCCTCTGCC GCCTTGTITT TGTCCGCGC ACAATCCGAA CGCGAGTGA
101 TGCCTGAGGT TTCTGCTGG CAGGAAAGA AAGGGGAAAA ACAGGCGGAG

```


151 CTGCGCTGAAA TCARAAGACGG TATGCCCGAT TTTCCCGAAC TTGCGCTGAT
 201 GCTTTTCGCAC GCGCTCAAAA CGGCAGTGTG TTAGCTGTGT GTCCGGTGTGG
 251 TCGCTTTCTCG CGGAAACTAT CTGCGCGCAG AATCCGAACC GGACAGGCCCC
 301 GTTCCGCCT..

5 This corresponds to the amino acid sequence <SEQ ID 488; ORF58>:

1 ..LRETAYVLDS FDRYFVVALA GLFFVRAQSE REWMREVSAA QEKKGKEQAE
 51 LPEIKDGMFD FPELALMLFH AVKAVYWLFG VGVVRFRCRNY LAHESEPDPR
 101 VPP..

Further work revealed the complete nucleotide sequence <SEQ ID 489>:

10 1 ATGTTTGGGA TAGTTTGGAT CGTTATTTTG TTGCTTGCGC TTGCGCGCTT
 51 GTTTTGTGTC CGCGCACAAAT CGAAGCGCGA GTGGATGCGC GAGGTTTCTG
 101 CGTGGCAGGA AAGAAAGGG GAAAAACAGG CGGAGCTGCC TGAAATCAAA
 151 GACGGTATGC CGGATTTTCC CGAACTTGCC CTGATGCTTT TCGATGCGGT
 201 CAAAACGGCA GTGATTGGC TGTTTGTGCG GTGCTGCGGT TTCTGCCGAA
 15 251 ACTATCTGCG GCACGAATCC GAAACGGACA GCGCCGTTCC GCGTGTCTCT
 301 GCRAACCGTG CGGATGTTCC GACCGCATCC GACGGATATT CAGACAGTGG
 351 AAACGGGACG GAAGAAGCGG AAGCGGAAGA AGCAGAAGCT GCGGAGGAAG
 401 AGGCTGCCGA TACGGAAGAC ATTGCAACTG CCGTAATCGA CAACCGCCGC
 451 ATCCCAITCG ACGGGAGTAT TCGTGAAGGG TTGATGCCGT CTCGAAGCGA
 20 501 AATTTTCGCG GTCCGTCGGG TTTTAAAGA AATCACTTGA GAAGAAGCA
 551 CGCGTGGTAA AACACGCGCG CCGTTAAGCG AAACGAAAAA ACGCTATATC
 601 GATGCATTTG AGAAAAACGA AACRCGCGTC OCCAAAGTCC GCGTGTGCGA
 651 TACCCCGATG GAAGGGCTGC AGATTATCGG TTTGGACGAC CCGTGTGCTG
 701 AAACGACGTA TTCCCATATG TTGATGTCGG GTCAAAAGAC GTTTTCCGAG
 25 751 TCTGCGGATT ACGGATTTTGA GCGGATTTT GAGAAGCAGC ATCCGTCTGC
 801 CTTTTCTGCA GTCAAAGCGG AAAATGCACG AATATGCGCG GTTCCACCGT
 851 ATGCGAGGCA GGGGAAAGGG CAGGCGGAGG CAAAATCCCC GGATGTTTCC
 901 CAAGGGCAGT CCGTTTCAGA GCGCACGGCC CTCGCGGATG CCGCGCGCGC
 951 CGTTTCCGTC AATTGGAAG AACCGAACAA GGCACCGGTT TCTCGGAGG
 30 1001 CGCGAATTTT TCGCCTGATT CCGGAAAGTC AGACGCTTGT CGGGAACAAGG
 1051 GATGTCGAAA TGCGCTCTGA AACCGAAAA GTTTTCACGG AAACCGTTTC
 1101 GTCTGTGGGA TACGGCGGTC CGGTTTATGA TGAACTGCC GATATCCATA
 1201 TTGAAGAACC TGCGCGGCCG GATGTTGGG TGGTCGAACC ACCCGAAGTG
 35 1251 CGGAAAGTTC CCATGACCGC AATCGATATT CAGCCGCGCG CTCGCGTATC
 1251 GGAAATCTAC AACCGTACCT ATGAACGCCG GTACGAGATC GAGCAGGTGC
 1301 AACGCAAGCG CATTGCGGAG ACCGACCATC TTGCGCATGA TGTTTTGAA
 1351 GAGAGTTTGG AGGAGAAAC CCGCGCTATT CGCGGTACCG CAGTGAAGG
 1401 TGCGGCAGAG CGGTCAAGCG GGCATATCTC TTCCGGAACC GAAGCGTTCG
 40 1451 GGCATGACAG TCAGGCGGTT TGTCCGTTTG AAAATGTGCC GTCTGAACCG
 1501 CCGTCTGCGC GGGTATCGGA TACGGAAGCG GATGAGGGGG CGTTCGCCAT
 1551 TGAAGAAGCC GGTGCGGTAT CGGACACCTC GCGGACAACC GACCTGTCTC
 1601 TGCTCTCGCT GTTCAATCCC GAGGCGCAGC AAACCGAAGA AGAATCTGTG
 1651 GAAAACAGCA TCACCATCGA AGAAAAATTG CGCGAGTTCA AAGTCAAGGT
 45 1701 CAAGGTTGTC GATTCTTATT CCGGCCCGGT AATTACGGGT TATGAATCG
 1751 AACCGGATGT CGGCGTGCGC GGCATTTCCG TTCTGAATCT GGAAAAAGAT
 1801 TTGGCGCGTT CGCTCGGCGT GGCATTCCAT CGCGTTGTGC AAACCATCCC
 1851 CGGCAAAACC TGCATGGGTT TGGAACTTCC GAACCGGAAA GCCCAATGA
 1901 TACGCTCGAG CGAAATCTTC AATTGCGCGG AGTTTGCAGA ATCCAAATCC
 50 1951 AAGCTGAGCG TCGCGCTCGG TCAGGACATC ACOSGACAGC CCGTGTGTAAC
 2001 CGACTTGGGA AAAGCAGCGG ATTATTTGGT TCGCGCGCAG ACGGTTTCGG
 2051 GCARATGATG GCGCTCTCAC GTGCTATGCT TGTCTATGCT TTTCAAGCT
 2101 GCGCGGAGAG ACGTGTGATG GATTGTGATG TGTCTGCAAA TGCTGGAATC
 2151 GAGCATTTAC GAAGGCATCC CGCACTGCTG GCCTCCGTCT GTTACCGATA
 2201 TGAAGCTGCG GCGAAACGCG CTGAAGTGTG GTGTTTAAAG AATGGAAAAA
 55 2251 CGCTACCGCC TGATGAGCTT TATGGCGGTG CGTAACTCTG CGGCTTCAAC
 2301 TCAAAAAATC GCGGAAGCGG CAGCAAGGGG AGAAAAATC GCGCAATCCG
 2351 TCAGCCTCAC GCGCGACGAT CCCGAACCTT TGGAAAAATC GCGGTTTATC
 2401 TTGGTCTGTT TCGATGAGTT TCGCGACCTG ATGATGACGG CAGGCAAGAA
 60 2451 AATCGAAGAA CTGATTGCCG GCGCTGCCCA AAAAGCCGCG GCGGACGCGA
 2501 TCCATTGAT TCTTGCCACA CAAGGCCCCA CGCTCGATGT CATCAGGGGT
 2551 CTGATTAAAG CGAACATCCC GAGCGGTATC GCGTTCCAAG TGTCCAGCAA
 2601 AATCGACAGC CGCACGATTCT TCGACCAAT GGGCGCGGAA AACCTGCTCG
 2651 TTCAGGCGGA TATGCTGTTT CTGCTGCCGG GTACTGCGTA TCCGCAAGCG
 2701 GTTCAAGGAT CATTGCTGTC GGAATGAAGG GTGACCGCGG CTCGCGAATA
 65 2751 TTTGAAGGAT CGGACTATGT TGCAGATATT TTGAGCGGG
 2801 CCGGCACGCA AGAGCTGCCG GGCATCGGCG GACGCGCGGA CGACGAGAAC

```

2851 GATCCGATGT ACGACGAGGC CGTATCCGTT GTCTCGAAA CGCGCAAAGC
2901 CAGCATTTTCG GGCCTACAGC AGCCCTTGGC TATCGGTAC AACCGCGCGG
2951 CGCGTCTGAT TGACAGATG GAGCGGAGAT GCATTGTGTC CGCACCGGAA
3001 CACAACGGCA ACCGTACGAT TCTCGTCCCG TTGACAATG CTGGA

```

5 This corresponds to the amino acid sequence <SEQ ID 490; ORF58-1>:

```

1 MFWIVLIVIL LLALAGLFFV RAQSEREWMR EVSAWQEKKG EKQALPEIK
51 DGMDDFPPELA LMLFHAVKTA VYWLFGVVVR FCNRYLAHES EDRPVPVPA
101 ANRADVPVTA DGVSDSGNGT EEAETEEAEA AEEBAADTED IATAVIDNR
151 IPFDRSIAEG LMPSESEISP VRPFVKEITL EATRALNSA ALRETKKRYI
201 DAFKNETAV PKVRVSDTPM EGLQIIGLDD PVLQRTYSHM FDADKEAFSE
251 SADYGFEFYP EKQHPSAFSA VKAENARNAP FHRHAGQGGK QAEAKSPDVS
301 QGQSVSDGTA VRDARRRVSV NLKEPNKATV SAEARISRLI PESQTVVGR
351 DVEMPSETEN VFTETVSSVG YGGFVYDETA DIHIEEPAP DAWVEPPEV
401 PKVEMTAIDI QPPFPVSEIY NRTYEPSPGS EQVQSRISAE TDHLADDVIN
15 401 GGWQETAAI ADDGSEGAEE RSSQYLSET EAFGHDSQAV CFFENVPSE
501 PSCRVSDETA DEGAFFSEET GAVSEHPLT DLLLPLPLNF EATQTEELL
551 ENSITIEEKL AEFKVVKVSV DSYSGPVITR YEIEPDVGVV GNSVINLEKD
601 LARSLGVASI RVVETIPGKT CMGLELFPNK RQMLRLSEIF NSFEFRESK
651 KLTLALGQDI TGQPVVTDLG KAPHLLVAGT TSGKSVGVN AMLSMLFKA
20 701 APEDVRMIMI DPKMLELSIY EGIPIHLLAP VTDMLAANA LWCVNMEEK
801 VVVDSFPELD WMTAKGKIEE LIARLAKAKR ANGLHILIAI QRSVDVITG
851 LTKANIPTRI AFQVSSKIDS RTILDDQMAE NLLQGGDMLF LLPGTAYPOR
901 VHGFASDEE VHRVVEYLKQ FGEEDYVDDI LSGGSEELP GIGRSQDDT
25 951 DPMYDEAVSV VLKTRKASIS GVQRALRIGY NRAARLIDQM EAGEIVSAPE
1001 HNGNRTILVP LDNA*

```

Computer analysis of this amino acid sequence predicts the indicated transmembrane region, and also gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

30 ORF58 shows 96.6% identity over a 89aa overlap with an ORF (ORF58a) from strain A of *N. meningitidis*:

```

              10      20      30      40      50      60
orf58.pep    LRETAYVLDSFDRYFVVALAGLFFVRAQSEREWMRVSAWQEKKGKQALPEIKDGM
35 orf58a     MFWIVLIVILLLALAGLFFVRAQSEREWMRVSAWQEKKGKQALPEIKDGM
              10      20      30      40      50
              70      80      90      100
orf58.pep    FPPELALMLFHAVKTAVYWLFGVVVRCNRYLAHESDPVP
40 orf58a     FPPELALMLFHAVKTAVYWLFGVVVRCNRYLAHESDPVPANRADVP
              60      70      80      90      100      110

```

The complete length ORF58a nucleotide sequence <SEQ ID 491> is:

```

45 1 ATGTTTGGTA TAGTTTGTAT CGTTATTTTG TTGCTTGCGC TTGCGCGCTT
51 GTTTTTTGTC CGCGCACAAAT CGAACCGGCA GTGATGCGC GAGGTTTCTG
101 GTTGGCGAGGA AAGAAAGAGG GAAAAACAGG CGGAGCTGCC TGAAATCAAA
151 GACGGTATGCG CCGATTTTCC CGAACTTGCC CTGATGCTTT TCATGCGCGT
201 CAACACGGGCA GTGATTGGC TGTATTGTCG TGTGCTCCGT TCTGCGCGAA
251 ACTATCGGCG GCACGAAATC GAACCGGACA GACCGCTCCG GCCTGCTCTC
301 GCAAAATCGTG CGGATGTTCC GACCGCATCC GACGATATT CAGACAGTGG
351 AAACCGGACG GAAGAAGCGG AAACCGAAGA ACAGAACTCG CGCGAGGANG
401 AGGCTGCCGA TACGGAAGAC ATTGCAACTG CCGTAATCGA CAACCGCGCC
451 ATCCCATTCG ACCGGAGTAT TGCTGAAGGG TTGATGCGGT CTGAAGCGGA
501 AATTTCGCCCG GTCCGTCGGG TTTTAAAGGA AATCACTTTG GAAGAAGCAA
551 CGCGTGCTTT AAACAGCGCG GCTTTAAGGG AAACGAAAAA ACGCTATATC
601 GATGCATTGT AGAAAAACGA AACAGCGGTC CCCAAGTCCG CGGTGTCGGA
651 TACCCGATG GAAGGGCTGC AGATTATCGG TTTGGACGAC CCTGTGCTTC
701 AACGCACGTA TTCGCGATG TTGATGCGG ACAAGAAGC GTTTTCCGAG

```

751 TCTGGCGATT ACGGATTTGA GCCGTATTTT AGAAGGAGC ATCCGTCGTGC
 801 CTTTTCGTCA GTCAAGCCG AAAATGCACG GAATGCGCCG TCCGCGCGTC
 851 ATCGAGGCGA GGGNAAGAGG CAGCGGGAGG CNAATCCCC GGATGTTTCC
 901 CANGGCCACT CCGTTTCAGA CGGCACAGCC GTCCGCGATC CNGCGCGCGC
 951 CGTTCCTCGTC AATTTGAAAG AACCGAACAA GCAACGGTT TCTGGGAGG
 1001 CGCGGATTTC GCGCTGATT CCGGAAAGTC GGACGGTTGT CGGGAAACGG
 1051 GATGTCGAAA TGCCGTCTGA AACCGAAAT GTTTTCACGG AAANTGTTTC
 1101 GTCTGTGGGA TACGCGCGTC CGGTTTATGA TGAACCTGCC GATATCCATA
 1151 TTGAAGAACC TGCCGCGCCC GATGCTTGGG TGCTCGAACC ACCCGAAGTG
 1201 CCGAAGTTC CCATGCCCCG AATNGATATT CCGCGCCGCG CTCGCCGTATC
 1251 GGAATCTAC AACCGTACCT ATGAACCGCC GCGAGGATTC GAGCAGGTGC
 1301 AACCGAGCCG CATTCGCCAA ACAGATCATC TTGCGGATGA TCTTTTGAAT
 1351 GGAGGTTGGC AGGAGGAAAC CGCCGCTATT GCGAATGAOG GCAGTGAGGG
 1401 TGTGGCAGAG OGGTCAAGCG GGCATATTTT GTCGAAAAOC GAAGCTTCG
 1451 GGCAATGACAG TCAGGCGGTTT TGTCGGTTTG AAAATGTGCC GTCTCAACGC
 1501 CCGTCCCAGCC GGGCATNGGA TACGGAAGCG GATGAAGGGG GTCTCCATC
 1551 TGAAGAAACC GGTGCGGTAT CCGAACCTCT GCGGCAACG GACCTGCTTC
 1601 TGCCGCGCCT GTTCAATCCC GCGCGACGCG AACCCGAAGA AGANCTGTTG
 1651 GNAACAGCA TCACCATCGA AGAAAAATNG CGCGGATCA AAGTCAAGGT
 1701 CAGAGTCTCT GATTCTTATT CCGCGCCCGT GATTACGGCT TATGAATCTG
 1751 AACCCGATGT CCGCGTCGCG GGCATTCGCG TTCTAAATCT GGAAAAAGAN
 1801 TTGGCGCGTT CGCTCGCGCT GCGTTCACCT CGCGTGTGCG AAACCTTCCT
 1851 CGGCAAAACC TGTATGGGTT TGAACCTTCC GACCCCGAAA CGCCAAATGA
 1901 TAGCCTGAG GAAATCTTC AATTGCGCCG AGTTTCCGGA ATCCAATCTC
 1951 AAGCTGACGC TCGCGCTCGG TCAGGACATC ACCGGACAGC CCGTCTTAAC
 2001 CGACTTGGCG AAGACACCGC ATTTTGTGGT TGCCGCGACG ACCGGTTCCG
 2051 GCAATCTCGT GGGTGTCAAC GCGATGATTC TGCTATGCT TTTCAAGGCC
 2101 GCGCCGGAAG ACGTGCATAT GATTATGATC GATCCGAAAA TGCTGAATTT
 2151 GAGCATTTCAC GAAGGCATCC CGCACTGCT CGCCCTGTTC GTTACCGATA
 2201 TGAAGCTGGC GGCAACGCG CTGAACGTGT GTGTAAAGAA AATGAAAAAA
 2251 CGCTACCGCC TGATGAGCTT TATGGGGTGT GCGAATCTTG CGGCTNICA
 2301 TCAAAAATC CCGCAAGCCG CAGCAAGGGG GCGAARAAT GCGAACCGGT
 2351 TCAGCTCTAC GCGCGCATC TCGAARACT TCGAARAAT GCGCTTTATC
 2401 TTGGTCTGTT TGATGAGTCT TCGCGACCTC ATGATGACGC CAGCGAAGAA
 2451 AATCGAAGA CTGATTGCCC GCCTCGGCCA AAAAGCCCGC CAGCGAGCGA
 2501 TCACTCTTAT CTTTGCACCA CAACGCCCCA GTGTTCATGT CATCAAGGGT
 2551 CTGATTAAAG CGAACATCCC GACGCGTATC GCGTTCACAG TGTCACGAA
 2601 AATCGACAGC CGCACGATTC TTGACCAAT GGGTGCAGAA AACCTGCTCG
 2651 GGCAGGGCGA TATGCTGTTT CTGCGCGCGG FACGCGCTA TCGCGAGCG
 2701 GTTCACGGCG CGTTTGCCCT GGATGAAGAG GTGACCGCGG TGCTCGAATA
 2751 TCTGAAACAG TTTGGCGAAC CGGACTATGT TGACGATATN TTAGCGCGCG
 2801 GTATGTCGGA CGATTTGCTG GGAATCAGCC GGAGCGCGGA CGCGGAACGC
 2851 GATCCGATGT ACGACGAGGC CGTCTCNGTT GTTTTGAATA CGCGCAAAAGC
 2901 CAGCATTTCT GCGGTGACAG GCGCATTCGG TATCGGCTAT AATCGCGCGC
 2951 CGCGTCTGAT TGACCAAGATG GAGCGGGAAG GCATTGTGTC CGACCGGAAA
 3001 CACAACGGCA ACGTACGAT TCTCGTCCCC TTNGACAATG CTGGA

This encodes a protein having amino acid sequence <SEQ ID 492>:

1 MFWIVLIVIL LLAAGLFFV RAQSEFWNR EVSAWQEKKG EKQAELEPEK
 51 DGMDFPEELA LMLEHAYTKA VYVLFVGVVR FCNRYLAHES EPDRFVFPAS
 101 ANRADVPFAS DQYSDSGMGT EEAETEEAEA EABRADTED IATRVINNR
 151 IFFDRSIAEG LMPSESEISP VRPVFKEITL EATRRLNLSA ALRTEKKRYI
 201 DAFKNETAVP PKVRVSDTFPM EQLQIGLDD PVLQRTYSRM FADKAEAFSE
 251 SADYGFPEFY EKQHFSAFSA VKAENARNAP FRRHAGQGGK QAEAKSPDVS
 301 QGQSVSDGTA VRDAXRRVSV NLKEPNKATV SAEARTSLRI PESRTVVGKR
 351 DVEMPESETEN VETEXVSSVG YGXVPYDETA DHTEEPAPAP wDAWVVEPEPV
 401 PKVPMFAXDI PPPPPVSEIY NRTYEPPAGF EQVGRSRIAE TDHLADDVIL
 451 GGWQEETAAL ANDGSEGVAE RSSQYLVSET EAFGHDSQAV CFFENVPESR
 501 PSRRAXDTEA DEGAFOSEET GAVSEHLPTT DLLLPLELNP GATQTEEXLL
 551 XNSITIEEKK AEFKVKVKV SVSYSGPVITR YEIEPVDGVR GNSVLNLEKX
 601 LARSLGVASI RVVETLLGKT CMGLELPNPK RQMLRLEIF NSPEFASKS
 651 KLTLALGQDI TGQPVVTDLG KAPHLVAGT TGSGKSVGVN AMILSMLFKA
 701 APEDVRHMLI DFMELLSIY ECIPIHLLAPV VTDMKLAAA LMKVNMWK
 751 RYRLMSFQVY RNLAGNKIKI ABAASGSKI GNPFSLTFDN PEPLKLEPEI
 801 VVVVDFADL YMTAGKKEIE LIARLAQKAR AAGIHLILAT QRSVVDVITG
 851 LIKANIPTRI AFQVSKSIDS RTIILDMQAE LMLCGGDMLF LPFGTAYPQR
 901 VHGAFAFDEE VHRVVEYLKO FGEPDYVDX LSGGMSDDL GISRSGDGT
 951 DFMVDEAVSV VLKTRKASIS GVQRLRIGY NNAARLIDQM EABGIVSAFE
 1001 HNGNRTILVP XDNA*

ORF58a and ORF58-1 show 96.6% identity in 1014 aa overlap:

		10	20	30	40	50	60
	orf58a.pep	MFWIVLIVILLALAGLFFVRAQSEREWREVSAWQEKKEQAELEPEIKDGMDFPFELA					
5	orf58-1	MFWIVLIVILLALAGLFFVRAQSEREWREVSAWQEKKEQAELEPEIKDGMDFPFELA					
		10	20	30	40	50	60
	orf58a.pep	70	80	90	100	110	120
10	orf58-1	LMLFHAVKTAVYWLFGVGVRFRCNRYLAHESEPPDRPVPPASANRAVDPTASDGYSDSGNGT					
		70	80	90	100	110	120
	orf58a.pep	130	140	150	160	170	180
15	orf58-1	EEAEETEEAAEEEAADTEDIATAVIDNRRIPFDRSIAEGLMPSESEISFVPVPVFKETIL					
		130	140	150	160	170	180
20	orf58a.pep	190	200	210	220	230	240
	orf58-1	EEATRALNSAALRETKKRYIDAFEKNETAVPKVRVSDTPMEGLQIILGLDDPVLQRTYSRM					
25		190	200	210	220	230	240
	orf58a.pep	250	260	270	280	290	300
30	orf58-1	FDADKEAFSESADYGFEPYFEKQHPSAFSAVKAENARNAPFRRHAGQCKGQAEAKSPDVS					
		250	260	270	280	290	300
35	orf58a.pep	310	320	330	340	350	360
	orf58-1	QGQVSDGTAVRDAXRRVSVNLKEPNKATVSAEARI SRLIPESRTVVGKRVDVEMPSETEN					
		310	320	330	340	350	360
40	orf58a.pep	370	380	390	400	410	420
	orf58-1	VFTXVSSVGYGXPFVYDETADIIIEEPAADFADWVVEPEVEKVPMPAXDIPPPPVSEIY					
		370	380	390	400	410	420
45	orf58a.pep	430	440	450	460	470	480
	orf58-1	NRTYEPFAGFEQVQRRIAEADHLADDVLNGWQEETAAIANDGSEGAERSSGQYLSET					
		430	440	450	460	470	480
50	orf58a.pep	490	500	510	520	530	540
	orf58-1	EAFGHDSQAVCPFFENVSPSRRAKDTEADGAFQSEETGAVSEHLPTDILLPLPLFNP					
55		490	500	510	520	530	540
	orf58a.pep	550	560	570	580	590	600
60	orf58-1	GATQTEEXLLXNSITIEEKXAEFKVKVKVDSYSGPVITRYEIEPDVGVRGNSVLNLEKX					
		550	560	570	580	590	600
	orf58a.pep	610	620	630	640	650	660
65	orf58-1	LARSLGVASIRVVETILGKTCMGLLPNPKRQMRILSEIFNSPEFAESKSKLTALAGQDI					
		610	620	630	640	650	660
		670	680	690	700	710	720

orf58a.pep	TGQPVVTDLGKAPHLLVAGTTGSGKSVGVNAMILSMLFKAAPEDVRMIMIDPKMLELSIY				
orf58-1	TGQPVVTDLGKAPHLLVAGTTGSGKSVGVNAMILSMLFKAAPEDVRMIMIDPKMLELSIY	670	680	690	700
		710	720		
orf58a.pep	EGIPHLLAPVVTDMKLAANALNWCNEMEKRYRLMSFMGVRNLAGKNQKIAEAAAARGEKI	730	740	750	760
orf58-1	EGIPHLLAPVVTDMKLAANALNWCNEMEKRYRLMSFMGVRNLAGFPNQKIAEAAAARGEKI	770	780		
orf58a.pep	GNPFSLTDPNPEPLKXLPFIIVVVDEFADIMMTAGKKIEELIARLAQKARAAGIHILILAT	790	800	810	820
orf58-1	GNPFSLTDPDPEPLEKLPFIIVVVDEFADIMMTAGKKIEELIARLAQKARAAGIHILILAT	830	840		
orf58a.pep	QRPSVDVITGLIKANIPTRIAFQVSSKIDSRITLDQMGAEENLLGGGDMFLFLPFGTAYPQR	850	860	870	880
orf58-1	QRPSVDVITGLIKANIPTRIAFQVSSKIDSRITLDQMGAEENLLGGGDMFLFLPFGTAYPQR	890	900		
orf58a.pep	VHGAFASDEEVHRVVEYLKQFGEPPYVDDXLSGGMSDDLLGISRSGGGETDPMYDEAVSV	910	920	930	940
orf58-1	VHGAFASDEEVHRVVEYLKQFGEPPYVDDILSGGSEELPGIGRSGDDTDPMPYDEAVSV	950	960		
orf58a.pep	VLKTRKASISGVQRLRIGYNRAARLIDQMEAEIGVSAPPEHNGNRTILVFXDNAX	970	980	990	1000
orf58-1	VLKTRKASISGVQRLRIGYNRAARLIDQMEAEIGVSAPPEHNGNRTILVPLDNAX	1010			

Homology with a predicted ORF from *N.gonorrhoeae*

ORF58 shows complete identity over a 9aa overlap with a predicted ORF (ORF58ng) from *N. gonorrhoeae*:

orf58.pep	ALMLFHAVKTAVVYVLFVGVVFRNRYLAHESEDPDRVPP	103
orf58ng	SEPDRVPVFASANRADVFTASDGYSDSGNG	30

The ORF58ng nucleotide sequence <SEQ ID 493> is predicted to encode a protein having partial amino acid sequence <SEQ ID 494>:

1	..SEPDRVPVPPA	SANRADVPTA	SDGYSDSGNG	TEEAETAAE	AAEEEAADTE
5	DIATAVIDNR	RIPFDRSIAE	GLMQSEKTS	FVRPVFKEIT	LEATALLS
10	AALREYKKRY	IDAEKNGTA	VPKYRVSDTP	MEGLQIIGLD	DVPLQRTYBR
15	MEADREKAFS	ESADYGFEPY	FEKQHPASFS	AVKARNARNA	PFRHQAQEK
20	GOAEAKSPDV	SGQGSVSDGT	AVRDARRRVS	VNLKEPNKAT	VSAAERISRL
25	IPESRTVVVK	RDVEMPSETE	NVFTETVSSV	GGGEPVVDRA	ADIHIEBPAA
30	PDAMVVEPPE	VPEVAVPEID	ILPPPVVSEI	YNRTYEPGAP	FEQAQRGRIA
35	ETDHLAADVL	NGGWQEETAA	IADDSSEGAA	ERSSGQYLSE	TEAFGHSQQA
40	VCPEFEDVSE	RPSCRVSDETE	ADEGAFQSE	TGAVSEHLPT	TDLLLPPLFN
45	PEATQTEEEL	LENSITIEEK	LAEFKVKVKV	VDSYSGPVIT	RYEIEPPDVG
50	RGNSVLNLEK	DLARSLGVAS	IRVVETIPGK	TCMGLLEPNP	KQIMIRLSEI
55	FNSPFAESK	SKLTALGQD	ITGQPVVTDL	GKAPHLLVAG	TTGSGKSVGV
60	NAMILSMLFK	AAPEDVRMIM	IDPKMLELSI	YEGITHLLAP	VVTDMKLAAN
65	ALNWCVNEME	KRYRLMSFMG	VRLNLAGFNQK	IAEAAAARGEK	IGNPFSLTDP
70	DPEPLEKLPF	IVVVVDEFAD	LMMTAGKKIE	ELIARLAQKA	RAAGIHILILA
75	QRPSVDVIT	GLIKANIPTR	IAFQVSSKID	SRTILDQMGGA	ENLLGGGDMFL
80	FLPFGTAYPQ	RVHGAFASDE	EVHRVVEYLK	QFGEPPYVDD	ILSGGSEEL
85	FGIGRSQDGE	TDPMYDEAVS	VVLKTRKASI	SGVQRLRIG	YNRAARLIDQ
90	MEAEIGVSAP	EHNGNRTILV	PLDNA*		

This partial gonococcal sequence contains a predicted transmembrane region and a predicted ATP/GTP-binding site motif A (P-loop; double underlined). Furthermore, it has a domain homologous to the FtsK cell division protein of *E. coli*. Alignment of ORF58ng and FtsK (accession number p46889) show a 65 % amino acid identity in 459 overlap:

5	ORF58ng: 467	IEEKLAEPFKVKVVDYSYSGFVITRYEIEPFDVGRNSVLNLEKDLARSLGVASIRVVET 526
	+E +LA+ <u>F</u> + <u>K</u> VV+ GPVITR+E+ GV+ +L +DLARSL ++RVVE	
	FtsK: 868	VEARLADFRKADVVNYSFGFVITRFEFLNLAGVKAARISNLSRDLARSLSTVAVRVVEV 927
10	ORF58ng: 527	IPGKTCMGLLEPNPKRMIRLSEIFNSPEFAESKSKLTALGQDITGQPVVTDLGKAPHL 586
	IPGK +GLELPN KRQ + L E+ ++ +F ++ S LT+ LG+D L G+PVV DL K PHL	
	FtsK: 928	IPGKPYVGLLEPNKKRQTVLYREVLDNAKFRDNPSPLTVVLGKDAGEPVVADLAKMPLH 987
15	ORF58ng: 587	LVAGTTGSGKSVGVNAMILSMLFKAAPEDVVRMIMIDPKMLELSIYEGITHLLAPVVTDMK 646
	LVAGTTGSGKSVGVNAMILSML+KA PEDVR IMIDPKMLELS+YEGI HLL VVTDMK	
	FtsK: 988	LVAGTTGSGKSVGVNAMILSMLYKAQPDVRFIMIDPKMLELSVYEGIPHLLTEVVTDMK 1047
20	ORF58ng: 647	LAANALNWCNEMEKRYRLMSFMGVRNLAGFNQKIAEAAARGEKIGNPFSLTPDDPEP--- 704
	AANAL WCNVME+RY+LMS +GVRNLG+N+KIAEA I ++ D +A	
	FtsK: 1048	DAANALRWCNVMEERRYKLMSALGVRNLGAGYNEKIAEADRMRRFIPDPYKPGDSMDAQH 1107
25	ORF58ng: 705	--LEKLPPFIVVVDEFADLMMTAGKKIEELIARLAQKARAAGIHLIATQRPSPVDVITGL 762
	L+K P+IVV+VDEFADLMMT GKK+EEELIARLAQKARAAGIHL+LATQRPSPVDVITGL	
	FtsK: 1108	PVLKEPYIIVLVDEFADLMMTVGKKVEELIARLAQKARAAGIHLVLATQRPSPVDVITGL 1167
30	ORF58ng: 763	IKANIPTRIAFQVSSKIDSRTILDQMGAEHLGGDMLFLPFGTAYQRVHGAFASDEEV 822
	IKANIPTRIAF VSSKIDSRTILDQ GAE+LLG GDM L+ P + P RVHGAF D+E V	
	FtsK: 1168	IKANIPTRIAFTVSSKIDSRTILDQAGAESLGMGDMLYSGPNSLTVPRVHGAFVRDQEV 1227
35	ORF58ng: 823	HRVVEYLKQFGE PDYVD DILSGGSGEELPGIGRSGDGETDPMDYEAHSVVLKTRKASISG 882
	HVV+ K G P YVD I S SE G G G E DP++D+AV V + RKASISG	
	FtsK: 1228	HVVQDKWARGRPQYVDIGTISDESSEGGAG-GFDGAEELDFLDQAVQVFTEKKRASISG 1286
40	ORF58ng: 883	VQRALRIGYNRAARLIDQMEAGIGVSAPEHNGNRTILVP 921
	VQR RIGYNRAAR+I+QMEA+GIVS HNGNR +L P	
	FtsK: 1287	VQRQFRIGYNRAARLIEQMEAQGIIVSEQHNGNREVLAP 1325

Further work on ORF58ng revealed the complete gonococcal DNA sequence to be <SEQ ID 495>:

1	ATGTTTGGGA TAGTTTGGAT CGTTATTgtg TTGCTTGCGC TTGCCGCGCT
5	GTTTTTTGTG CGCGCACAAT CGCAACGCGA GTGGATGCGC GAGGTTTCTG
10	CGTGGCAGGA AAGAAAGGG GAAAAACAGG CGGAGCTGCC TGAATCAAA
15	GACGGTATGC CGGATTTCOC CGAGTTTTCO CTGATGCTTT TCCATGCCGT
20	CAAAACGGCA GTGATTGGC TGTTTGTGCG TGTGTCGGT TTCTGCCGAA
25	ACTATCTGGC GCACGAATCC GAACCGGACA GGCCGTTTCC GCTGCTTCT
30	GCAACCGCTG CGGATGTTCC GACCGCATCC GACGGGTATT CAGACAGTGG
35	AAACGGGACG GAAGAAGCGG AAACGGAAGC AGCAGAAGCT GCGGAGGAAG
40	AGGCTCCGCA TACGGAAGAC ATTGCAACTG CCGTATTCGA CAACCGCCGC
45	ATCCcatTCG ACCGGAGTAT TCGTGAAGGG TTGATGCAAT CTGAAGACAA
50	AACTCTGCCO CTTGCTGCCG TTTTAAAGGA AATCACTTTG GAAGAAGCAA
55	CGCGTGTCTT AAGCAGCGCG GCTTTAAGGG AAACGAAAAA ACGCTATATC
60	GATGCTATTG AGAAAAAGCG AACAGCGCTG CCAAAAGTAC GCGGTCTCGA
65	TACCCCGATG GACGCGCTGC AGATTATTCG TTTGGACGAC CCGTGTCCGA
70	AGCGACAGTA TTCCGCTATG TTTGATGCGG ACACAAGAGC GTTCTCCGAG
75	CTTCGCGATT ACGGATTTGA GCGGTATTT GAGAAGCAGC ATCCGTCTGC
80	CTTTTCTGCA GTCAAAGCGG AAATGCAAGC GAATGCGCGC TTCCGCGCTC
85	ATGCGAGGCA GGAGAAAGGG CAGGCGGAGG CAATAACCCC GGATGTTTCC
90	CAAGGGCAGT CCGTTTCAGA GGGCACAGCC GTCCCGGATG CCGCGCGCGC
95	CGTTTCCGCT AATTGTAAAG AACCGAACAA GGCAACGGTT TCTGCGGAGG
100	CGCGGATTTC GCGCGTGATT CCGGAAAGTC GGACGGTTGT CGGGAACAGG
105	GATGTCGAAA TGCGCTCTGA AACCGAAAAT GTTTTCAAGC AAACCGCTTT
110	GTCTGTGGGA TACGGCGGTC CGGTTTATGA TGAAGCTGCC GATATCCATA
115	TGAAGAGGCC TGCGCGCGCC GATGCTTGGG TGGTCGAACC ACCCGAAGT
120	CCGAGGTTAG CCGTACCGCA AATCGATATT CTGCGCGTATC CTTCCGATAT
125	GGAAATCTAC AACCGTACCT ATGAGCCGCG GCGAGGATTC GAGCAGGCG
130	AACGACGCGC CATTGCGGAA ACGGACCATC TTGCGCGTGA GTTTTGAAT

1351 GGAGGTTGGC AGGAGGAAAC CGCGCTATT GCAGATGACG GCAGTGAGGG
 1401 TCGCGCAGAG CGGTCAAGCG GGCAATATCT GTCGGAAACC GAAGCGTTTCG
 1451 GGCATGACAG TCAGGCGGTTT TGTCCGTTTG AGAATGTGCC GTCTGAACGCG
 1501 CGGTCTGCGC GGTGATCGGA TACGAGAGCG GATGAGAGGG CTTTCACATC
 1551 GGAAGABACC GGTGCGGTAT CGGACACACT CGGACAAAC GACCTGCTGC
 1601 TGCTCCGCT GTTCAATCCC GAGGCGAGCG AAACCGAAGA AGAATGTTTG
 1651 GAAACACGCA TCACCATCGA AGAAAAATTG GCGGAGTTCA AAGTCAAGGT
 1701 CAAGGTTGTC GATTCTTTATT CCGGCCCGCT GATTACCGGT TATGAAATCG
 1751 AACCOCATGT CGGCGTGCGC GGCAATTCCG TTCTGATTTT GGA AAAAGAC
 1801 TTGGCGCGTT CGCTCGGCGT GGCTTCCATC CGGTTGTGCG AAACCATCCC
 1901 CGGCAAAACC TGCATGGGTT TGGAACTTCC GAACCGGAAA GCGCAATGTA
 1951 TACGCTGAG CGAAATTTTC AATTGCGCCG AGTTTGCCGA ATCCAAATCC
 2001 AAGCTGACGC TCGCGCTCGG TCAGGACATT ACCGGACAGC CGTCGTATAC
 2051 CGACTTGGCG AAAGCACGCG ATTTGCTGGT TGCCGGCAGC ACCGTTTCGG
 2051 GCAAAATCGT GGGTGTCAAC GOGATGATTC TGTCTATGCT TTTCAAAGCC
 2101 GCGCGGGAAG ACGTGCATAT GATTATGATC GATCCGAAAA TGCTGGGATT
 2151 GAGCATTTAC GAAGGCATCA CGACCTGCTC GCGCCCTGTC GTTACCGATA
 2201 TGAAGCTGSC GGCAACGCGC CTGAACCTGGT GTGTTAAAGA AATGGA AAAA
 2251 CGCTACCGCC TGATGAGCTT TATGCGCGTG CGCAATTTTG CGGCTTTCAA
 2301 CCAAAAATTC CGCGAAGCCG CAGCAAGSGG AGCAAACTCGT GGCRAATCGT
 2351 TCAGCCTCAC CCGCGACGAT CCGGAACCTT TGGAAAATCT GCGCTTTATC
 2401 GTGGTCTGGT TCGATGAGTT GCGGATTTG ATGATGACGG CAGGCAAGAA
 2451 AATCGAAGAA CTGATTGCGC GCTTCGCCCA AAAAGCCCGC GCGCGAGCA
 2501 TCCACCTTAT CCTTGCCACA CAACGCCCCA GCGTCGATGT CATCACGGGT
 2551 CTGATTAAAG CGAACATCCC GACGCGTATC GCGTTCGAAG TGTCAGGAAA
 2601 AATCGACAGC CGCACGATTC TCGACCAAAAT GCGCGCGGAA AACCTGCTCG
 2651 GTACGGGCGA TATGCTGTTT CTGCGCGCGG GTACTGCCTA TCCGACGCGC
 2701 GTTCAAGCGC CGTTTGCTCT GGGATGAAGG GTGACACGCG TGTGCGAATA
 2751 TCTGAAGCAG TTGCGCGAGC CGGACTATGT TGACGATATT TTGAGCGCGC
 2801 GCGGACGCGA AGAGCTGCCC GGCATCGGGC GTGACGCGGA CGGCGAAAC
 2851 GATCCGATGT ACAGACGAGC CGTATCGGTT GTCTCGAAAA CGCGCAAAAC
 2901 CAGCATTCGT GGCCTACAGC AGCGCTTGCG CATCGCTATC AACCGCGCGC
 2951 CGCGCTGAT TGCACAAATG GAAGCGAAG GCATTGTGTC CGCACCGGAA
 3001 CACAAACGCA ACCGTACGAT TCTCGTCCCC TTGGCAATG CTTGA

This corresponds to the amino acid sequence <SEQ ID 496; ORF58ng-1>:

1 MFWIVLIVIV LLAALAGLFFV RAQSERENMR EVSAWQEKKG EKQAEIPEIK
 51 DGMPPDFPEFS LMLFHAVKTA VYWLFGVGVV FCRNYLAHES EPPDRVPFPA
 101 ANRADVPFAS DGYSDSGNGT EEAETEAEEA AEEAAADTED IATAVIDNRR
 151 IPFDRSIAEG IMQSESKTSP VRPVFKEITL EEAATRALSSA ALRETKKRYI
 201 DAFKKNGTAV PKVRVSDTPM EQLQIIGLDL PVLQRTYSRM FDADKEAFSE
 251 SADYGFEPYF EKQHPSAFSA VKAENARNAP FRRHAGQKKG QAEAKSPDVS
 301 QGQSVSDGTA VRDARRRVSV NLKEPNKATV SAEARISRLI PESRTVVGKR
 351 DVEMFSETEN VFETVTSSVG YGGFVYDEAA DIHIEEPAAP DAWVVEPEV
 401 FEVAVPEIDI LPFPFVSEIY NRTYEPFAGF EQARQSRIAE TDHLADVLN
 451 GGWQESTAAI ADDGSEGAEE RSSGQYLSET EARGHDSQAV CPFEDVPSE
 501 FSCRVSDETA DEGAFOSEET GAVSEHLPTT DLLLPFLFNP EATQTEEBLL
 551 ENSITTEEKL AEKFKVYKVV DSYSGGVTVR VYIEPDPGVGR GNSVNLMEKD
 601 LARSLGVASI RVVETTPGKT CMGLELNPVK RMQIRLSIEIF NSPEFAESKS
 651 KLTLLALGQDI TGQFVVVDLG KAPHLVAGT TGSCKSGVGV AMILSMPLTA
 701 APEDVRMIMI DPKMLELSIY EGITHLLAPV VTMKLAANA LNCWVNEKE
 751 RYRLMSFMGV RNLAGFNQKI AEAARGEKI GNPFLSLTPD FEPELEKLPFI
 801 VVVVDEFADL MMTAGKKEIE LIARLAQKAR AAGIHLILAT QRPVSDVITG
 851 LKANIPETRI AFQVSSKIDS RTILDQMGAE NLLGQGDMLF LPPTAYTAPR
 901 VHGFASDEE VHRVVEYLKQ FGEPDYVDI LSGGSEELP GIGRSGDGET
 951 DPMYDEAVSV VLKTRKASIS GVQRALRIGY NRRARLIDQM EAEGIVSAPE
 1001 HGNRTILTV LDNA*

ORF58ng-1 and ORF58-1 show 97.2% identity in 1014 aa overlap:

	10	20	30	40	50	60
orf58-1.pep	MFWIVLIVILLALAGLFFVRAQSERENMR	EVSAWQEKKG	EKQAEIPEIK	KDGMPPDFPELA		
orf58ng-1	MFWIVLIVILLALAGLFFVRAQSERENMR	EVSAWQEKKG	EKQAEIPEIK	KDGMPPDFPEFS		
	10	20	30	40	50	60
	70	80	90	100	110	120
orf58-1.pep	LMLFHAVKTAVYWLFGVGVVFCRNYLAHES	EPDRVPFPA	ANRADVPFAS	DGYSDSGNGT		

	orf58ng-1	LMLFHAVKTAVYWLFGVGVFRFCRNYLAHESEPDVRVPPASANRADVPTASDGYSDSGNGT	70	80	90	100	110	120
5	orf58-1.pep	EEAETEEAAAEAAADTEDIATAVDNRRIPFDRSIAEGLMPSESEISFVRPVFKETTL	130	140	150	160	170	180
	orf58ng-1	EEAETEEAAAEAAADTEDIATAVDNRRIPFDRSIAEGLMQSESKTSPVRPVFKETTL	130	140	150	160	170	180
10	orf58-1.pep	EEATRALNSAALRETKKRYIDAFEKNETAVPKVRVSDTPMEGLQIIGLDDPVLQRTYSHM	190	200	210	220	230	240
	orf58ng-1	EEATRALSSAALRETKKRYIDAFEKNGTAVPKVRVSDTPMEGLQIIGLDDPVLQRTYSRM	190	200	210	220	230	240
15	orf58-1.pep	FDADKEAFSESADYGFPEYFEKQHPSAFSAVKAENARNAPFRRHAGQKQAEAKSPDVS	250	260	270	280	290	300
20	orf58ng-1	FDADKEAFSESADYGFPEYFEKQHPSAFSAVKAENARNAPFRRHAGQKQAEAKSPDVS	250	260	270	280	290	300
25	orf58-1.pep	QGQSVSDGTAVRDARRVSVNLKEFNKATVSAEARSRLIPESQTVVGKRDVEMPSETEN	310	320	330	340	350	360
	orf58ng-1	QGQSVSDGTAVRDARRVSVNLKEFNKATVSAEARSRLIPESRTVVGKRDVEMPSETEN	310	320	330	340	350	360
30	orf58-1.pep	VFTETVSSVGGYGGPVYDETADIIHIEEPAAPDAWVVEPPEVKVPMTAIDIQPPPVSEIY	370	380	390	400	410	420
	orf58ng-1	VFTETVSSVGGYGGPVYDEAADIIHIEEPAAPDAWVVEPPEVFEVAPVPEIDILPPPVSEIY	370	380	390	400	410	420
35	orf58-1.pep	NRTYEPSPGFEQVQVRSRIAETDHLADVDVNGGWQETAAIADDGSEGAERSSGQYLSET	430	440	450	460	470	480
	orf58ng-1	NRTYEPGAGFEQAQVRSRIAETDHLADVDVNGGWQETAAIADDGSEGAERSSGQYLSET	430	440	450	460	470	480
40	orf58-1.pep	EAFGHDSQAVCFENVPSERPSCRVSDDTEADGAFPEETGAVSEHLPTDLLPLPLFNP	490	500	510	520	530	540
	orf58ng-1	EAFGHDSQAVCFEDVVPSERPSCRVSDDTEADGAFQSEETGAVSEHLPTDLLPLPLFNP	490	500	510	520	530	540
45	orf58-1.pep	EATQTEEELENSITIEKLAEPKVKVVDVSYSGPVITRYEIEPDVGVGRNSVLNLEKD	550	560	570	580	590	600
50	orf58ng-1	EATQTEEELENSITIEKLAEPKVKVVDVSYSGPVITRYEIEPDVGVGRNSVLNLEKD	550	560	570	580	590	600
55	orf58-1.pep	LARSLGVASIRVVETIPGKTCMGLELPNPKRQMRILSEIFNSPEFAESKSKLTLAGQDI	610	620	630	640	650	660
	orf58ng-1	LARSLGVASIRVVETIPGKTCMGLELPNPKRQMRILSEIFNSPEFAESKSKLTLAGQDI	610	620	630	640	650	660
60	orf58-1.pep	TGQPVVTDLGKAPHLLVAGTTGSGKSVGVNAMILSMLFKAAPEDVRMIMIDPKMLELSIY	670	680	690	700	710	720
	orf58ng-1	TGQPVVTDLGKAPHLLVAGTTGSGKSVGVNAMILSMLFKAAPEDVRMIMIDPKMLELSIY	670	680	690	700	710	720
65	orf58-1.pep	EGIPHLLAPVVTDMKLAANALNWCNVNEMEKRYRLMSFMGVNRNLAFGNOKIAEAAARGEKI	730	740	750	760	770	780
	orf58ng-1	EGITHLLAPVVTDMKLAANALNWCNVNEMEKRYRLMSFMGVNRNLAFGNOKIAEAAARGEKI	730	740	750	760	770	780
70			790	800	810	820	830	840

	orf58-1.pep	GNPFSITPDDPEPELEKLPFFIVVVVDFEADLMMTAGKKIEELIARIQAQARAAGIHLILAT	
	orf58ng-1	GNPFSITPDDPEPELEKLPFFIVVVVDFEADLMMTAGKKIEELIARIQAQARAAGIHLILAT	
5		790 800 810 820 830 840	
	orf58-1.pep	QRPSVDVITGLIKANIPTRIAFQVSSKIDSRITLDMQGAENLLGGQDMLFLLPGTAYPOR	
	orf58ng-1	QRPSVDVITGLIKANIPTRIAFQVSSKIDSRITLDMQGAENLLGGQDMLFLLPGTAYPOR	
10		850 860 870 880 890 900	
	orf58-1.pep	VHGAFASDEEVRHVRVYLYLKQFGEPDYVDIISGGGSEELPGIGRSGDDETDPMYDEAVSV	
	orf58ng-1	VHGAFASDEEVRHVRVYLYLKQFGEPDYVDIISGGGSEELPGIGRSGDDETDPMYDEAVSV	
15		910 920 930 940 950 960	
	orf58-1.pep	VHGAFASDEEVRHVRVYLYLKQFGEPDYVDIISGGGSEELPGIGRSGDDETDPMYDEAVSV	
	orf58ng-1	VHGAFASDEEVRHVRVYLYLKQFGEPDYVDIISGGGSEELPGIGRSGDDETDPMYDEAVSV	
		910 920 930 940 950 960	
20	orf58-1.pep	VLKTRKASISGVQRALRIGYNRAARLIDQMEAGIVSAPEHNGNRTLIVLPLNAX	
	orf58ng-1	VLKTRKASISGVQRALRIGYNRAARLIDQMEAGIVSAPEHNGNRTLIVLPLNAX	
		970 980 990 1000 1010	
	orf58-1.pep	VLKTRKASISGVQRALRIGYNRAARLIDQMEAGIVSAPEHNGNRTLIVLPLNAX	
	orf58ng-1	VLKTRKASISGVQRALRIGYNRAARLIDQMEAGIVSAPEHNGNRTLIVLPLNAX	
		970 980 990 1000 1010	

Furthermore, ORF58ng-1 shows significant homology to the *E.coli* protein FtsK:

```

25  sp|P46889|PTSK_ECOLI_CELL_DIVISION_PROTEIN_PTSK >gi|1651412|gnl|PID|d1015290 (D1
    division protein FtsK [Escherichia coli] >gi|1651418|gnl|PID|d1015296 (D90727) Cell
    division protein FtsK [Escherichia coli] >gi|1787117 (AE000191) cell division
    protein FtsK [Escherichia coli] Length = 1329
    Score = 576 bits (1469), Expect = e-163
    Identites = 301/459 (65%), Positives = 353/459 (76%), Gaps = 5/459 (1%)

30  Query: 556  IEELAEFKVKVVKVDSYSGPVITREYIEFDVGVGRNSVLNLEKDLARSLGVASIRRVET 615
    +E +LA+F++K VV+ GVITR++L+ GV+ + NL +DLARSL ++RVVE
    Sbjct: 868  VEARLAFRIKADVYNSPSGPEVITRFEFLNLAFGPKAARISLNRDLARSLTAVVRVVE 927

35  Query: 616  IPGKTCMGLELNPKNRQIMRLSEIFNSPEFAESKSKLTALGQDDITGPQVVDLKGAPHL 675
    IPGK +GLELPN KRQ + L +E ++ ++ + S LT+ IG+DI G+PVP DL K PHL
    Sbjct: 928  IPGKGYGLELNNKQKRVITREVLNNAKFRDSEFLTVLGVKDIGEAPVADLAKMPEH 987

40  Query: 676  LVAGTTSGSGKSVGVNAMLSMLKFAEPDVRVIMIDPKMKLELSYIEGITHLLAPVVTDMK 735
    LVAGTTSGSGKSVGVNAMLSMLK+K PEDVR IMIDPKMKLELS+YEGI HLL VVTDMK
    Sbjct: 988  LVAGTTSGSGKSVGVNAMLSMLKYAQPEDVRFIMIDPKMKLELSYIEGITHLLVVTDMK 1047

    Query: 736  LAANALNWCNEMEMERYKLSMFGVNRNLAGNKGKIEAAARGEKINPESLTPDDEP-- 793
    AANAL WCNEME+RY+LMS +GVRLNLA+KXIEA I+P+ D+
    Sbjct: 1048  DAANALRWCNEMEMERYKLSMGAIGVRLNLAGYNEKIEAADRMRRPIPDYPYWKPGDSMDAQH 1107

    Query: 794  --LEKLPFVVVVVDEFADLMATGAKKIEELLARLAQKARAAGIHLILATQRPSVDVITGL 851
    L+K P+IVV+V+V+DEFADLMTT GKK+ELLARLAQKARAAGIHL+LATQRPSVDVITGL
    Sbjct: 1108  FVLKKEPYVVLV+VDEFADLMMTVGKVEELLARLAQKARAAGIHLVLATQRPSVDVITGL 1167

50  Query: 852  IKANIPTRIAFGVSSKSIDSRITLDQGAENLLGMDMLFLPPGTAYPQRVHGAFASDEVE 911
    IKANIPTRIAF VSSKSIDSRITLDQAE+LGG DML+ P + P RVHGAF +D+V
    Sbjct: 1168  IKANIPTRIAFTAVSSKSIDSRITLDQAGAESLLGMDMLYSGFNSTLFEVRVHGAFVRQDEV 1227

55  Query: 912  HRVVEYLKQCFEPYVDYDLISGGSGSEELFGIRSGDGETPMYDEAVSVVLKTRKASIG 971
    H VV+ K G P YVD I S SE G G E DP++D+AV V +RKASIG
    Sbjct: 1228  HAVVQWKARGRPQYVDYDITSDSESGGAG-GFDGAELTFLDQAVQVFTVEKKASIG 1286

    Query: 972  VORALRIGYNRAARLIDQMEAGVIGSAPBHGNGRTLP 1010
    VQR RIGYNRAAR+I+QMEA+GIVS HNGNR +L P
    Sbjct: 1287  VQORFRIGYNRAARIEQMEAAGVIGSECHGNGRNLVAP 1325

```

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 59

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 497>:

```

1   ATGATTTCAT  AAGAAACCT  CATCAAGAA  CTCTCTTTTA  CCGCGCTGG
5   51  CATTTTCGTC  GTCCCTCTGG  CGGTATTGGT  CTCACGCGAC  GCAATCAACC
101 TGCTCGGCGC  TGCCGCCGAC  GGGC...GTGA  TCGCCATCGA  TGCCGTTGT
151 GCATTGGTGC  GCTTCTGGGT  C.....

901 .....A  TTGCCATCGG  TTTGTTTTTA  ATTTACACAA  ACGGGCTGAC
951 CCTGCTTTTT  GAAGCCGTGG  AAGACGGCAA  AATCCATTTT  TGGCTCGGAC
10  1001 TGCTGCGTAT  GCACATTATC  ATGTTTGTCC  TTGCACCTCAT  CCTGTTGGCG
1051 GTCCGCGAGT  TGCCCGACCA  GCCCTTCTGG  CAGGCGGTTG  GCAAAAGTCT
1101 GACATTGAAA  GCGGGAATAA  GA

```

This corresponds to the amino acid sequence <SEQ ID 498; ORF101>:

```

1   MIYQRNLIKE  LSFTAVGIFV  VLLAVLVSTQ  AINLLGRAAD  GXVIAIDAVL
15  51  ALVGFVW...  .....
      //
301 ...IAIGLFL  IYQNGITLLF  EAVEDGKIHF  WLGLLEPMHI  MFVLAILLRL
351 VRSMPSQPFW  QAVGKSLTLK  GGR*

```

Further work revealed the complete nucleotide sequence <SEQ ID 499>:

```

20  1   ATGATTTCAT  AAGAAACCT  CATCAAGAA  CTCTCTTTTA  CCGCGCTGG
51  51  CATTTTCGTC  GTCCCTCTGG  CGGTATTGGT  CTCACGCGAC  GCAATCAACC
101 TGCTCGGCGC  TGCCGCCGAC  GGGC...GTGA  TCGCCATCGA  TGCCGTTGT
151 TTGCTCGGCT  TCTGGTCAT  CGGTATGACG  CGGCTTTTGC  TGCTGTTGAC
25  201 CGCATTTCAT  AGTACGTTGA  CGGTGTTGAC  CCGCTACTGG  CCGCACAGCG
251 AARTGTCCGT  CTGGCTATCC  TCGCGATTGG  CATTTGAACA  ATGGATAGCG
301 CCGGTGATGC  AGTTTGCCGT  GCCGTTTGCC  GTTTTGGTTG  CCGTCATGCA
351 GCTTTGGGTG  ATACCGTGGG  CAGAGCTACG  CAGCGCGGAA  TACGCTGAAA
401 TCTTGAAGCA  GAAGCGAGAA  TTGCTTTTGG  TGGAGGCGAG  CGAGTTCAAC
30  451 AGTTTGGGCA  AGCGCAACGG  CAGGGTTTAT  TTGTGCGAAA  CTTTGCATAC
501 CGAATCCGCG  ATCATGAAAA  ACCTGTTTCT  GCGCGAAGAG  GACAAAAACG
551 GCGGCGACAA  CATCATCTTC  GCCAAGAGAG  GTAACCTCTC  GCTGAACGAC
601 AACAAACGCA  CGCTCGAATT  GCGCCACGGC  TACGTTTACA  GCGGCACGCC
651 CCGACGCGCC  GACTACAATC  AGGTTTTCCT  CCAAAAACTC  AACCTGATTA
701 TCAGCACACAC  GCCCAATCTC  ATCGACCCCG  TTTCCACCGC  CCGTACCAAT
35  751 CCGACGCGCC  AATCGATTGG  CAGCAGCAAC  CCGCAACATC  AGCGCGAATT
801 GATGTGGGCG  ATCTCGCTGA  CCGTCAGCGT  CCTCTACTCT  TGCTGCTGTC
851 CCGTGC CGCT  TTCTTATTTC  AACCGCGGCA  GCGGACATAC  CTACAATATC
901 TTGATTGCCA  TCGGTTTGTG  TTTAATTATC  CAAACCGGGC  TGACCTGCTC
951 TTTTGAAGCC  GTGGAAAGCG  GCAAAATCCA  TTTTGGGCTC  GGACTGCTGC
40  1001 CTATGCACAT  TATCATGTGT  CGCGTTGAC  TCATCCTGTT  GCGGCTCGCG
1051 AGTATGCCCA  GCCAGCCCTT  CTGGCAGGCG  GTTGCAAAAA  GTCTGACATT
1101 GAAAGCGGGA  AAATGA

```

This corresponds to the amino acid sequence <SEQ ID 500; ORF101-1>:

```

45  1   MIYQRNLIKE  LSFTAVGIFV  VLLAVLVSTQ  AINLLGRAAD  GRVAIDAVLA
51  51  LVGFVWVGMT  PLLLVLTAFI  STLTVLTRYW  RDSEMSVWL  CGLALKQWIR
101 FVMQFAVEFA  VLVAVMQLWV  IPWAEILRSR  YAEILKQKQR  LSLVEAGEFN
151 SLKRWGRVY  EVTETDETS  IMKNLFLRQ  DKNGSDNIF  ARGNFSLND
201 NKRTLLELHG  YRYSGTFPRA  DYNQSVGFI  NLLSTTHKI  LDPVSHRRI
251 PTAQLIGSSN  PQHAEILMWR  ISLTVSVLL  LLAVELSYF  NFRSHYNI
50  301 LIAIGLFLIY  QNGITLLF  EVDGKIHF  WLGLLEPMHI  MFVLAILLVR
351 SMPSQPFQWA  VGKSLTLKGG  K*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF101 shows 91.2% identity over a 57aa overlap and 95.7% identity over a 69aa overlap with an ORF (ORF101a) from strain A of *N. meningitidis*:

		10	20	30	40	50
orf101.pep		MIYQRNLIKELSFTAVGIFVLLAVLAVSTQAINLLGRAADGXVIAIDAVLALVGFVWX				
5	orf101a	MIYQRNLIKELSFTAVGIFVLLAVLAVSTQAINLLGXADXRK-AIDAVLALVGFVXXM				
		10	20	30	40	50
				//		
				90	100	110
10	orf101.pepIAIGLFLIYQNGLTLLFEAVEDGKIHFWLGL				
	orf101a	LTVSVLLLCLLAVPLSYFNPRSGHTYNILXAIIGLFLIYQNGLTLLFEAVEDGKIHFWLGL				
		280	290	300	310	320
						330
		120	130	140	150	
15	orf101.pep	LPMHIIMFVLAILLLVRSMPSQPFVQAVGKSLTLKGGKX				
	orf101a	LPMHIIMFVIAIVLLVRSMPSQPFVQAVGKSLTLKGGKX				
		340	350	360	370	

The complete length ORF101a nucleotide sequence <SEQ ID 501> is:

	1	ATGATTATC	AAGAAACCT	CATCAAAGAA	CTCTCTTTTA	CCGCCGTCGG
20	51	CATTTTCGTC	GTCTCTCTGG	CGGTATTGCT	CTCCAGCGAG	CAATCAACC
	101	TGCTCGGCCN	TGCCGCCGAC	NGCGTNTCCG	CATCGATGCG	CGTGTGGCA
	111	TTGGTCGGCT	TCTGGGTCNN	NGNATGACG	CCGCTTTTGT	TNGTGTGAC
25	201	CGCATTTATC	AGTACGTGA	CCGTGTTGAC	CCGCTACTGG	CNGACACGG
	251	AAATGTCGGT	CTGNTATCC	TGCGGATTGG	CATTGAACA	ATGGATACG
	301	CCGCTGATGC	AGTTTGGCT	GCCGTTTCCG	GTTTTGGTGT	CCGTATGCA
	351	GCTTTGGGTG	ATACCGTGG	CAGAGCTACG	CAGCGCGGAA	TACGCTGAAA
30	401	TCCTGAAGCA	GAAGCAGGAA	TTGTCTTTGG	TGAGGACGAG	CGGTTCAAC
	451	AGTTTGGGCA	AGCGCAACGG	CAGGCTTTAT	TTTGTGAAA	CCTTCGATAC
	501	CGAATCCGCG	ATCATGAAA	ACCTGTTCTC	CGCGGACAG	GACAAAAAG
	551	GCGCGCAGCA	CATCATCTTC	NCCAAAGAAA	GTAACTTCTC	CGTGAACAG
	601	AACAAACGCA	CGCTCGAATT	GCGCCACGCG	TACCGTTACA	GCGGCACGCC
	651	CGGACGCGCC	GACTCAATC	AGGTTTCCCT	CNAAAACCTC	AACCTGATTA
35	701	TCAGCACAC	GCACAACTC	ATCGACCGCG	TTTCCGACCG	CTGACNATN
	751	CCNACNGGCC	ACTGATTTG	CAGCAGCAC	CCGCACTACT	ANGCGGAAT
	801	GATGTGGCGC	ATCTCGCTGA	CCGTACGCGT	CCCTCTACTC	TGCTCGCTG
	851	CCGTGCGCGT	TTCCTATTTC	AACCGCGCA	GCGGACATAC	CTACRATATC
	901	TTGANTGCCA	TCGGTTTGT	TTTAATTAC	CAGAACGGCG	TGACCTCGCT
	951	TTTGAAGCC	GTGGAAGAC	GCAAAATCCA	TTTTGGCTC	GGACTGCTGC
40	1001	CTATGCACAT	CATCATGTTT	GTCATCGCAA	TGCTACTTCT	GCGCGTCGCG
	1051	AGCATGCCCA	GCCAGCCCTT	CTGGCAGCG	GTTGGCAAAA	GTCGTACATT
	1101	GAAAGCGGGA	AARTGA			

This encodes a protein having amino acid sequence <SEQ ID 502>:

	1	MIYQRNLIKE	LSFTAVGIFV	VLLAVLAVSTQ	AINLLGXAAD	XRKXDAVLA
45	51	LVGFVWXXMT	PLLLVLTAFT	STLTVLTRYW	RDSEMSVWXS	CGLALKQWR
	101	FVMQFAVFFA	VLVAVMLQWV	IPWAEILRSRE	YAEILKQQR	LSLVEAGGFN
	151	SLGRNGRVYF	VETFTDTS	IMKNLFLREQ	DRNGGDNIF	XKESNFSLND
	201	NKRTLELHGH	YRYSGTTCRA	DYNQVSEFKI	NLLISTEKL	TDPSVHRRTX
	251	PTAQLTGSSN	QKHAEIMWR	ISLTVSVLLI	CLLAVPLSYF	NPRSGHTYWI
50	301	LXAIIGLFLIY	QNGLTLLFEA	VEDGKIHFWL	GLLPMHIIMF	VIAIVLLVR
	351	SMPSQFFWQA	VGKSLTLKGG	K*		

ORF101a and ORF101-1 show 95.4% identity in 371 aa overlap:

	orf101a.pep	MIYQRNLIKELSFTAVGIFVLLAVLAVSTQAINLLGXADXRKXDAVLA	60
55	orf101-1	MIYQRNLIKELSFTAVGIFVLLAVLAVSTQAINLLGRAADGXVIAIDAVLALVGFVWVIGMT	60
	orf101a.pep	PLLLVLTAFTSTLTVLTRYWRDSEMSVWXSGLALKQWIRPVMQFVAVPFAVLVAVMLQWV	120
	orf101-1	PLLLVLTAFTSTLTVLTRYWRDSEMSVWLSGLALKQWIRPVMQFVAVPFAVLVAVMLQWV	120
60	orf101a.pep	IPWAEILRSREYAEILKQKQELSLVEAGGFNSLGRNGRVYFVETFTDTS	180
	orf101-1	IPWAEILRSREYAEILKQKQELSLVEAGGFNSLGRNGRVYFVETFTDTS	180

	orf101a.pep	DKNGGDNIIFXKESNFSLNDNKRITELRHGYRYSGTPTGRADYNQVSFXKLNLIISTTPKL	240
	orf101-1	DKNGGDNIIFAKEGNFSLNDNKRITELRHGYRYSGTPTGRADYNQVSFQKLNLIISTTPKL	240
5	orf101a.pep	IDPVSHRRRTXPTAQLIGSSNPQHXAEIWMRISLTVSVLLCLLAVPLSYFNPRSGHTYNI	300
	orf101-1	IDPVSHRRRTIPTAQLIGSSNPQHXAEIWMRISLTVSVLLCLLAVPLSYFNPRSGHTYNI	300
10	orf101a.pep	LXAIGFLFIYQNGLTLLFEAVEDGKIHFHWGLLEPMHIIMFVIAIVLLRVRSMPSQPFWQA	360
	orf101-1	LIAIGFLFIYQNGLTLLFEAVEDGKIHFHWGLLEPMHIIMFAVALLLRVRSMPSQPFWQA	360
15	orf101a.pep	VGKSLTLKGGK	371
	orf101-1	VGKSLTLKGGK	371

Homology with a predicted ORF from *N.gonorrhoeae*

ORF101 shows 96.5 % identity in 57aa overlap at the N-terminal domain and 95.1% identity in 61aa overlap at the C-terminal domain, respectively, with a predicted ORF (ORF101ng) from *N. gonorrhoeae*:

20	orf101.pep	MIYQRNLIKELSFYAVGIFVVLAVLVSTQAINLLGRAADGXVIAIDAVLALVGFWV	57
	orf101ng	MIYQRNLIKELSFYAVGIFVVLAVLVSTQAINLLGRAADGRV-AIDAVLALVGFWVIGM	59
25		//	
	orf101.pep	IAIGFLFIYQNGLTLLFEAVEDGKIHFHWGL	333
	orf101ng	SLTVSVLLCLLAVPLSYFNPRSGHTYNIILIAIGFLFIYQNGLTLLFEAVEDGKIHFHWGL	331
30	orf101.pep	LLEPMHIIMFVALILLRVRSMPSQPFWQAVGKSLTLKGGK	373
	orf101ng	LLEPMHIIMFVIAIVLLRVRSMPSQPFWQAVG	362

The ORF101ng nucleotide sequence <SEQ ID 503> is predicted to encode a protein having partial amino acid sequence <SEQ ID 504>:

35	1	MIYQRNLIKE	LSFTAVGIFV	VLLAVLVSTQ	AINLLGRAAD	GRVAIDAVLA
	1	LVGFVWIGMT	PLLLVLTAFT	SLTVLTRYW	RDSEMSVWLS	CGLALKQWIR
	101	PVMQFAVFPA	ILIAVMQLVW	IPWAEIWSRE	YAEILKQKQE	LSLVEAGEFN
40	151	NLGRNNGRVY	FVETFDTEG	IMKNLFLREQ	DKNGGDNIIF	AKEGNFSLKD
	201	NKRITELRHG	YRYSGTPTGR	DYNQVSFQKL	NLIISTTPKL	IDPVSHRRTI
	251	STAQLIGSSN	POHAEIWMR	ISLTVSVLL	CLLAVPLSYF	NPRSGHTYNI
	301	LIAIGFLFIY	QNGLTLLFEA	VEDGKIHFWL	GLLPMHIIMF	VIAIVLLRVR
	351	SMPSQPFWQA	VG...			

Further work revealed the complete nucleotide sequence <SEQ ID 505>:

45	1	ATGATTTCAT	AAAGAAACCT	CATCAAGAA	CTCTCTITTA	CCGCGTCCG
	15	CATTTCGCT	GTCCCTCTTG	CGGTCTTCTG	GTCCACGAC	GCATCAACC
	101	TGCTTGGCCG	CGCAGCTGAC	GGCGGTGTG	CCATCGATGC	CGGTGTGCC
	151	TTAGTCGCT	TCTGGGTCAT	CGGTATGACC	CCGCTTTTTC	TGCTGTGAC
50	201	CGCATTCATC	AGCAGCTGTA	CGGTATTGAC	CCGCTACTCG	CGCGACAGCG
	251	AAATGTCGGT	CTGGCTATCC	TGCGGATTGG	CGTGTAAACA	GCGGATACGC
	301	CCGCTCATGC	AGTTTGCCGT	GCCGTTTGCC	ATCCTGATTG	CCGTCATGCA
	351	GCTTTGGGTG	ATACCGTGGG	CAGAGCTGCG	CAGCCGCGAA	TATGCCGAAA
	401	TTTGAAGCA	GAAGCAGGAA	TTGCTTTTGG	TGGAAGCCGG	CGAGTTCAAT
55	451	AACCTGGGCA	AGCGCAACGG	CagggtttaT	TtcgtcgaaA	CGTTTGACAC
	501	GgaatccgCG	ATCATGAAAA	ACCTGTTcct	GcCGCAACAG	GACAAAACAG
	551	gcggcgacaA	CATCATCTTC	GCcaaaGaaq	gtaactTctc	gctgaaggaC
	601	AACAAAGcaA	cgctcgaaTT	CGCGCAACGG	TACCGTTACA	CGCGcaagcC
	651	CGaaCGGc	gctcgaATC	AgGTtTcctt	cCRAAAactC	aaactgATTa
	701	TCAGCACCAC	GCCCAAAcTT	ATCGaccCGG	TTTCCACC	CGCACCATT

5 751 tgcacCGCCCC AActGATTGG CAGCAGCAAT CCGCAACATC AGGCAGAAAT
 801 GATGTGGCGC ATCTCGTGGA CCGTCAGCGT CCTCCTGCTC TGCCTACTCG
 851 CCGTGC CGCT TTCTATTTC AACCCGCGCA GCGGACATAC CTACCAATATC
 901 TTGATTGCCA TCGGTTTGT TTTAATTAC CAARACGGGC TGACCCCTGCT
 951 TTTTCAGACC GTGGAGAGAC GCAAAATCCA TTTTGGCTC GGACTGCTGC
 1001 CTATGCACAT CATCATGTC GTCATGCCA TCGTACTTCT GCGCGTCGCG
 1051 AGTATGCCCA GCCAGCCCTT CTGGCAGGCG GTTGGCAAAA GTCTGACATT
 1101 GAAAGGcgGA AATGA

This corresponds to the amino acid sequence <SEQ ID 506; ORF101ng-1>:

10 1 MIYQRNLKE LSFTAVGIFV VLLAVLVSTQ AINLLGRAAD GRVAIDAVLA
 51 LVGFVWIGMT PLLLVLTAFI STLTVLTRYW RDSEMSVWLS CGIALKQWIR
 101 FVMQFAVPFA ILIAVLMQWV IPWAELRSRE YAEILKQKQE LSLVEAGEFN
 151 NLGKRNGRVY FVETFDTESE IMKNLFLREQ DKNGGDNIIF AKEGNFSLKO
 201 NKRTLELRHG YRYSGTPGRA DYNQVSFQKL NLIISTTPKL IDPVSHRRTI
 251 STAQLIGSSN PQHQAEIMWR ISLTVSVLLL CLLAVPLSYF NFRSGHTYNI
 301 LIAIGLFLIY QNGLTLLFEA VEDGKIHFWL GLLPMHIIMF VIAIVLLVRV
 351 SMPSPQFWQA VGKSLTLKGG K*

ORF101ng-1 and ORF101-1 show 97.6% identity in 371 aa overlap:

20 orf101-1.pep 10 20 30 40 50 60
 orf101ng-1 MIYQRNLKELSF
 10 20 30 40 50 60
 25 orf101-1.pep 70 80 90 100 110 120
 orf101ng-1 PLLLVLTAFIST
 70 80 90 100 110 120
 30 orf101-1.pep 130 140 150 160 170 180
 orf101ng-1 IPWAELRSREYAE
 130 140 150 160 170 180
 35 orf101-1.pep 190 200 210 220 230 240
 orf101ng-1 DKNGGDNIIFAKE
 190 200 210 220 230 240
 40 orf101-1.pep 250 260 270 280 290 300
 orf101ng-1 IDPVSHRRTIPTA
 250 260 270 280 290 300
 45 orf101-1.pep 310 320 330 340 350 360
 orf101ng-1 LIAIGLFLIYQNG
 310 320 330 340 350 360
 50 orf101-1.pep 370
 orf101ng-1 VGKSLTLKGGKX
 370
 55 orf101-1.pep 370
 orf101ng-1 VGKSLTLKGGKX
 370

60 Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is

predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 60

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 507>:

```

5      1  ..GGTGGTGGTT TTATCAATGC TTCCTGTGCC ACTTTGACGA CAGCCAAACC
      51  GCAATATCAA GCAGGAGACC TTAGCGCTTT TAAGATAAGG CAAGGCAATG
     101  TTGTAAATCGC CGGACACGGT TTGGATGCAC GTGATACCGA TTACACACGT
     151  ATTCTCAGTT ATCATTCCAA AATCGATSCA CCGTATGGG GACAAAGTGT
     201  TCGTGTCTCT CGGGACACAA ACGATGTGSC CCGAACGCT GATCGACATT
     251  CGCGTATTCT CAATAATGCT GCTGCCAATA CGTCAAACAA TACAGCCAC
     301  AACGSCACAC ATATCCCTTT ATTTGCGATT GATACAGGCA AATTAGGAGG
     351  TAT.GTATGC CAACAAATC ACCTTGATCA GTACGTCGA GCAAGCAGGC
     401  ATTCGTAA
  
```

This corresponds to the amino acid sequence <SEQ ID 508; ORF113>:

```

15      1  ..GGGFINASCA TLTTAKPQYQ AGDLSAFKIR QGNVVIAGHG LDARDTDYTR
      51  ILSYHSKIDA FVWGQDVRVW AGQNDVAATG DAHSPILNNA AANTSNTAN
     101  NGTHIPLFAI DTGLGGXVC QQNHLDQYGR ASRHS*
  
```

Computer analysis of this amino acid sequence gave the following results:

Homology with with pspA putative secreted protein of *N.meningitidis* (accession AF030941)

20 ORF and pspA show 44% aa identity in 179aa overlap:

```

      orf113  GGGFINASCATLTTAKPQYQAGDLSAFKIRQGNVVIAGHGLDARDTDYTRILSYHSKIDA 60
      GGG INA+ TLT+ P G+L+ F+ G VVI G GLD D DYTRILS ++I+A
      pspA    GGGLINAASVTLTSGVPVLNNGNLGTGFDVSSGKVIIGKGGLTSDADYTRILSRAAEINA 256

      orf113  FVWGQDVRVWAGQNDVAATGDAHSPILXXXXXXXXXXXXXGTHIPLFAIDTGLGGMYA 120
      VMG+DV+VV+G+N + G + P AIDT LGGMYA
      pspA    GVWGKDVVVSGRKNLDFDG-----SLAKTASAPSSSDSVTPTV AIDTATLGGMYA 307

      orf113  NKITLISTVEQAGIRNQGFASAGNVAVNAEGKLVNTGMIATGENHAVSLHARNVHN 179
      +KITLIST A IRN+G+ FA+ G V ++A+GKL N+G I A ++ A+ V N
      pspA    DKITLISTNGAVIRNKGRIFAATGGVTLSDAGKLSNSGSIDAA----EITISAQTVDN 362
  
```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF113 shows 86.5% identity in 52aa overlap at the N- terminal part and 94.1% identity in 17aa

35 overlap at the C-terminal part with a predicted ORF (ORF113ng) from *N. gonorrhoeae*:

```

      orf113  GGGFINASCATLTTAKPQYQAGDLSAFKIR 30
      orf113ng SHPSQLNGYIEVGGRRAEVVIANPAGIAVNGGGFINASRATLTGQPYQAGDFSGFKIR 224

      orf113  QGNVVIAGHGLDARDTDYTRILSYHSKIDAPVWGQDVRVWAGQNDVAATGDAHSPILNNA 90
      orf113ng QGNVVIAGHGLDARDTDYTRILVCQQNHLDQYGRTSRHS 263

      orf113  IDTGKLGXVCQQNHLDQYGRASRHS 135
      orf113ng DFGFKIRQGNVVIAGHGLDARDTDYTRILVCQQNHLDQYGRTSRHS 263
  
```

The complete length ORF113ng nucleotide sequence <SEQ ID 509> is predicted to encode a protein having amino acid sequence <SEQ ID 510>:

1 MNKTLRYVIF NRKRGAVVAV AETTKREGKS CADSGSGSVY VKSVSFIFTH
 51 SKAFCSFALG FSLCLALGTV NIAFADGIIT DKAAPKTQQA TILQTGNIGIP
 101 QVNIQTPTSA GVSVNQYQAF DVGNRGAILN NSRSNTQTQL GSWIQGNFWL
 151 TRGEARVVVN QINSHPSQL NGYIEVGGRR AEVVIANPAG IAVNGGGFIN
 201 ASRATLTGTQ PQYQAGDFSG FKIRQNGNAVI AGHGLDARDT DFTRIILVQQQ
 251 NHLQYGTST RMS*

Based on this analysis, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 61

10 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 511>:

1 ..TCAACGGGAC ATAGCGAACA AAATTACACT TTGCCGCGAG AAATCACACG
 51 CAACATTTC A CTGGGTTTCAT TTGCCCTATGA ATGCGCATCG CAAGCATATA
 101 GGCATCATGC GCCACGCCAA GGCATGAGT TGCGCGAAG CAACGGTATA
 151 TCGCTACCT ATAGCTCCAA TTCTTTTACC CCATTACCCA GCAGCAGCTT
 201 ATACATTATC AATCCTGTCA ATAAAGGCTA TCTTGTGTAA ACGATCCAC
 251 GCTTTGCCAA CTACCGTCAA TGGTTGGGTA TGACTATAT GCTGGACAGC
 301 CTCACACTAG ACCCAACCAA TTACATATAA CGTTTGGGTG ATGGTTATTA
 351 CGAGCAAGCT TTAATCATGC ACACATACGC AGACGTGACA GGGCATCGTC
 401 GTTTAGACGG TTATCAAAC GACGAGAGAC AATTATTAAGC CTTAATCGAT
 451 AATGGCGCGA CTGCGGCACG TTGATGAAT CTCAGCGTTG GCATTCGATT
 501 AAGTGCCGAG CAAGTAGCGC AACTGACCAG CGATATTGTT TGGTTGGTAC
 551 AAAAAGAGT TAAGCTTCTT GATGGCGCA CACAAACCGT ATTGGTGCCA
 601 CAGGTTTATG TACGCGTTAA AAATGGCGAC ATGACCGGTA AAGGTGCATT
 651 GTTGTCAAGC AGCAATACAC AAATCAATGT TTCAGGCAGC CTGAAAACT
 701 CAGGCACGAT TGCAGGCGC AATGCGCTTA TTATCAATAC CATTACGCTA
 751 GACCAATATC GTGGGCGTAT TCATGCGCAA AAATCAGCGG TTACGCCAC
 801 ACACAGACAT AATAATATTG GCGCATGCT TTCTGCCGAA CAGACATTAT
 851 TGCTCAACGC AGGCAACAC ATCAACAGCC AAAGCACACC CGCCAGCAGT
 901 CAATAACAC AAGGCAGCAG CACTACATC GACCGAATG CAGGATTTTA
 951 TATCAGAGC AAAGAAAAAG GTGTTT..

This corresponds to the amino acid sequence <SEQ ID 512; ORF115>:

1 ..STGHSEQNYT LPREITRNI SLGSFAYESH KALSHHAPSQ GTELPQSNIGI
 51 SLPYTSNSFT PLPSSSLYII NPNVNGYIV TDFRFRANYRQ WLGSYMLDS
 101 LKLDPNNLHK RLGDGYEQR LINEQIAELT GHRRLDGYON DEEQKALMD
 151 NGATAARSMN LSVGIALSAE QVAQLTSDIV WLQKEVKLP DGGTQTVLTP
 201 QVYVRVKGND IDGKGALLSG SNTQINVS GS LKNSGTIAGR NALINTDNL
 251 DNIGRIHAQ KSAVTATQDI NNIGGMLSAE QTLILNAGNN INSQSTTASS
 301 QNTGSSSTYL DRMAGIYITG KEKGV..

Computer analysis of this amino acid sequence gave the following results:

40 Homology with the pspA putative secreted protein of *N.meningitidis* (accession number AF030941)

ORF115 and pspA protein show 50% aa identity in 325aa overlap:

Orf115: 1 STGHSEQNYTLPREITRNI SLGSFAYESH KALSHHAPSQ GTELPQSNIGI 60
 STG+S Y E++ +I +G AY+ + + P + + NGI +T
 pspA: 778 STGYRSRSPYEPAPSVS-SIRMGISAYKGYAPQASDIPGTGVVFAENGIRHPTFT----- 831
 Orf115: 61 PLPSSSLYIINPNVNGYIVETDPRFRANYRQWLGSYMLDSLKLDPNNLHKRLGDGYEQR 120
 LP+SSL+ I P NKGYL+ETDP F +YR+WLGS YML +L+ DPN++HKRLGDGYEQR+
 pspA: 832 -LPNSSLFAIAPNNKGYLIETDPAFTDYRKWLGSYMLAALQQQDPNNH IHKRLGDGYEQK 890
 Orf115: 121 LINEQIAELTGHRRLDGYONDDEEQKALMDNGATAARSMNLSVGIALSAEQVAQLTSDIV 180
 L+NEQIA+LTG+RRLDGY NDDEEQKALMDNG T A+ +L+ GIALSAEQVA+LTSDIV
 pspA: 891 LVNEQIAKLTGYRRRLDGYTNDDEEQKALMDNGITIAKELQTLFGIALSAEQVARTLSDIV 950

Orf115: 181 WL VQKEVKLPDGGTQTVLVPQVYVRVKNQDIDGKGALLSGSNTQINVSGLSKN-SGTIAG 239
 WL + V LPDG TQTVL P+VYVR + D++G+GALLSGS I SG+++N G IAG
 pspA: 951 WLENETVTLPDGTTQTVLKPKVYVRARPKDMNQGALLSGSVVDIG-SGAIENRGGLIAG 1009

5 Orf115: 240 RNALIINTDTLDNIGGRIHAQKSAVTATQDINNIGGMLSAEQTELLNAGXXXXXXX 299
 R ALI+N + N+ G + + A DI N G + AE LLL A
 pspA: 1010 REALILNAQNIKNLQSDLGKKNIFAAGSGDITNGS-IGAENALLKASNINIESRSETRS 1068

10 Orf115: 300 XXXXXXXXXYLDRMAGIYITGKEKG 324
 + R+AGIY+TG++ G
 pspA: 1069 NQNEQGSVRNIGRVAGIYLTGRQNG 1093

Homology with a predicted ORF from *N.gonorrhoeae*

ORF115 shows 91.9% identity over a 334aa overlap with a predicted ORF (ORF115ng) from

N.gonorrhoeae:

orf115.pep STGHSEQNYTLPREITRNISLGSFAYESHK 31
 orf115ng NEQTFGEKKVFSENGKLHNYWRARRKGHDGTGHRQNYTLPEEITRDISLGSFAYESHK 71

20 orf115.pep ALSHHAPSQGTETLPQSN-----GISLPTNSFTPLPSSSLYIINPVNKGYLVT 81
 orf115ng ALSRHAPSQGTETLPQSNRDNIRTAKSGISLPTPNSTPLPGSSLYIINPANKGYLVT 131

25 orf115.pep DPRFANYRQWLGSDYMLDSLKLDPNNLHKLRLGDGYEQRLINEQIAELTGHRRLDGYQND 141
 orf115ng DPRFANYRQWLGSDYMLGSLKLDPNNLHKLRLGDGYEQRLINEQIAELTGHRRLDGYQND 191

30 orf115.pep EEQFKALMDNGATAARSMNLSVGIALSAEQVAQLTSDIVWLQKEVKLPDGGTQTVLVPQ 201
 orf115ng EEQFKALMDNGATAARSMNLSVGIALSAEQVAQLTSDIVWLQKEVKLPDGGTQTVLVPQ 251

35 orf115.pep VYVRVKNQDIDGKGALLSGSNTQINVSGLSKNSGTIAGRNALIINTDTLDNIGGRIHAQK 261
 orf115ng VYVRVKNQDIDGKGALLSGSNTQINVSGLSKNSGTIAGRNALIINTDTLDNIGGRIHAQK 311

40 orf115.pep SAVTATQDINNIGGMLSAEQTELLNAGNNINSQSTASSNTQGSSTYLDRMAGIYITGK 321
 orf115ng SAVTATQDINNIGGILSAEQTELLNAGNNINNSQSTAKSSQNAQGSSTYLDRMAGIYITGK 371

orf115.pep EKGV 325
 orf115ng EKGVLAAQAGKDINIAGQISNQSDQGQTRLQAGRDLNLTQVQTKYQEIHFADNHTIR 431

An ORF115ng nucleotide sequence <SEQ ID 513> was predicted to encode a protein having amino acid sequence <SEQ ID 514>:

1 MLVQTEKDLG HNEQTFGEKK VFSENGKLH YWRARRKGHD ETGHRQNYT
 51 LPPEITRDIS LGSFAYESHK KALSRHAPSQ GTETLPQSNRD NIRTAKSNGI
 101 SLPTYPNSFT PLPGSSLYII NPANKGYLVE TDPFRFANYRQ WLGSDYMLGS
 151 LKLDPNNLHKL RLGDGYEQYR LINEQIAELT GHRRRLDGYQN DEEQFKALMD
 201 NGATAARSMN LSVGIALSAE QAAQLTSDIV WLQKEVKLP DGCTQTVLMP
 50 251 QVYVRVKNKG IDGKGALLSG SNTQINVSGL LKNSGTIAGR NALIINTDTL
 301 DNIIGGRIHAQ KSAVTATQDI NNIGGILSAE QTELLNAGNN INNQSTAKSS
 351 QNAQGSSTYL DRMAGIYITG KEGVLAQAQ GKDINIAGQ ISNQSDQGQT
 401 RLQAGRDINL DTVQTKGYQE IHFDADNHTI RGSSTNEVGSS IQTKGDVTLTLL
 55 451 SGNNLAKAAK FVGSAGKTLA VYAKNDITIS SGIHAGQVDD ASKHTGRSGG
 501 GNKLIVTDKA QSHHTAQSS TFEKGQVVLQ AGNDANILGS NVISINGTRI
 551 QAQNHVIRIGT TQTQSQSEYV HQQKSGSLMS AGIGETIGSK TNIQENQSQS
 601 NEHTGSTVGS LKQDITIVAS KHYEQTGSNV SSPEGNLIS TQSDMIGAAQ
 651 NQLNKSITQT YEQKGLTVAF SSPVTDLAQQ AIAVAHKAQ QFDKAKTTAL
 701 MPWRLPQMVG RLFKQAKPK K*

Further work revealed the following partial gonococcal DNA sequence <SEQ ID 515>:

1	TTGCTTGTGC	AAACAGAAAA	AGACGGTTTG	CATAACGAGC	AAACCTTTGG
51	CGAGAAGAAA	GTCTTCAGCG	AAATATGGTAA	GTTCACAAAC	TACTGGCGTG
101	CGCGTCGTAA	AGGACATGAT	GAACACAGGC	ATCGTGAACA	AAATATATAT
151	TTGCGGGAGG	AAATCACACG	CGACATTTCA	CTGGGTTTCA	TTGCCCTATGA
201	ATCGCATAGC	AAAGCATTAA	GCCCTCATGC	GCCCAGCCAA	GGCCTATGAT
251	TGCCCAACAG	TAAACCGGAT	AATATCCGTA	CTGCGAAAG	CACGCGTATT
301	TGCTACCTCT	ATACGCCCAA	TTCTTTTACC	CAATTACCCG	GCACAGCGTT
351	ATACATTATC	AATCCTGCCA	ATAAAGGCTA	TCCTGTGTAA	ACCGATCCAC
401	GCTTTGCCAA	CTACCGTCAA	TGGTTGGGTA	GTGATATAT	GCTGGGACGC
451	CTCAAACTAG	ACCCAAACAA	TTTACATAAA	CGTTTGGGTG	ATGGTTATTA
501	CGAGCAACGT	TTAATCAATG	AACAATCGC	AGAGCTGACA	GGGCTAGCTC
551	GTTTAGACGG	TTATCAAAAC	GACGAAGAAC	AAITTTAAAGC	CTTAATGGAT
601	AATGGCGCGA	CTGCGGCACG	TTGATGAAT	CTCAGCGTTG	CGATTGCATT
651	AAGTGGCCAG	CAAGCAGCGC	AACTGACGAC	CGATATTGTT	TGGTTGGTAC
701	AAAAAGAAGT	TAAACTTCTT	GATGGCGGCA	CACAAACCGT	ATTGATGCCA
751	CAGGTTTATG	TACGCGTTAA	AAATGGCGGC	ATAGACGGTA	AAGGTGCATT
801	GTTGTCAGGC	AGCAATACAC	AAATCAATGT	TTCAGGCAGC	CTGAAAAACT
851	CAGGCACGAT	TGCAGGGCGC	AATGCGCTTA	TTATCAATAC	CGATACGCTA
901	GACAATATCG	GTGGGCGTAT	TCATGCGCAA	AAATCAGCGG	TTACGGCGAC
951	ACAAGACATC	AATAAATATG	GCGGCAATCT	TTCTGCGGAA	CAGACATTAT
1001	TGCTCAATGC	GCGTAACACG	ATCAACAACG	AAACACAGGT	CAGACAGATG
1051	CAAAATGCAC	AAGGTAGCAG	CACCTAACCT	GACCAATGSC	CAGGTAATTA
1101	TATCACAGGC	AAAGAAAAAG	GTGTTTTAGC	AGGCCAGGCA	GGCAAAAGCA
1151	TCAACATCAT	TGCCCGTCAA	ATCAGCAATC	AATCAGATCA	AGGGCAAAAC
1201	CGGCTGCGAG	CAGGACGCGA	CATTAACTGT	GATACGGTAC	AAACCGGCAA
1251	ATATCAAGAA	ATCCATTTTG	ATGCCGATAA	CCATACCATC	CGAGGTTCAC
1301	CGAACGAAGT	CGGCAGCAGC	ATTCAAACAA	AGGGCGATGT	TACCCtatTG
1351	TCAGGGAATA	ATCTCAATGC	CAAAAGCTGC	GAACTCGGCA	GCGCAAAAGG
1401	CACACTTGCC	GTGTATGCTA	AAANTGACAT	TACTATCAGC	TCAGGCTATC
1451	ATGCCGGCCA	AGTTGATGAT	CGCTCCAAAC	ATCAGCGCAG	AAGCGCGCGC
1501	GGTAATAAAT	TAGTCAATTAC	CGATAAAGCC	CAAAAGTCAT	ACGAAACTCG
1551	TCAAAGCAGC	ACCTTTGAAG	CGAAGCAAGT	TGATATTGAC	GACGGAACAG
1601	ATGCGACATC	CTCTGGTATC	AACTGATATT	CCGATTAATG	CACCCGGATT
1651	CAGCAGAGCA	ATCATGTTTC	CAATTCATGC	ACCCCAATCT	AAGCCCAAGG
1701	CGAAACCTAT	CATCAAAACC	AAAAATCAGG	ATTGATGAGT	CGAGGTATGC
1751	GCTTCACATAT	TGGCAGCAAG	ACAAACACAC	AAGAAACCCA	ATCCCAAGAG
1801	AACGAACATA	CAGGCAGTAC	CGTAGGCAGC	CTGAAAGCGC	ATACCAACAT
1851	TGTTGCAAGC	AAACACTACG	AACAACACGG	GAGCAACGTT	TCCAGCCCTG
1901	AGGGCAACAA	CCATTATCAGC	ACGCAAAAGTA	TCCGATAATTG	CGCAGCACAA
1951	AACCAATTAA	ACAGCAAAAC	CACCCAAACC	TACGAACAAA	AAGGCTTAAC
2001	GGTGGCATTG	AGTTGCGCCG	TTACCGATTT	GGCACAACAA	CGGATTGCCG
2051	TAGCACACAA	AGCAGCAAC	AAGTCGAGCA	AGCAAAAAAC	GACCGCGTTA
2101	ATGCCATGGC	GGCTGCCAAT	GCAGGTTGGC	AGGCTCATCA	AACAGGCACAA
2151	GGCGCACAAA	ACTTAG			

45 This corresponds to the amino acid sequence <SEQ ID 516; ORF115ng-1>:

1	LLVQTEKEDGL	HNEQTFCEKK	VFSNGKLIIN	YWRARRKGHD	ETGHREQNYT
51	LFEEITRDIS	LGSFAYESHG	KALSRAPSPQ	GTFLPQSNRD	NIRTAKSNGI
101	SLPYTPNSPT	PLPGSSLYII	NPANKGYLVE	TDPRANVRO	WISGDMYMG
151	LKLDPMNLHK	RLGDGYEYQR	LINEQIAELT	GHRRLDGYQN	DEEQTKALMD
201	NGATAARSMN	LSVGIALSAR	AAQLTSDIV	WLVOKEVLKP	DGDTQTVLMP
251	QVYVRVKNKG	IDKGALLSG	SNTPQINVSGS	LKNSGTIAGR	NALIINTDTL
301	DNIGGRHIAQ	KSAVTATQDI	NNIGGILSAR	QTLLLNAGNN	INNOSTAKSS
351	QNAQGSSTYL	DRMAGIYITG	KEKGVLAAQA	GKDINIAGQ	ISNQSDQGGT
401	RLQAGRDLNL	DTVQTGKYQE	IHFADNHTI	RGSTNEVGS	IQTKGDTVLL
451	SGNNLNAKAA	EVGSAKGTLA	VYAKNDITIS	SGTHAGVDD	ASKTGTGRSGG
501	GKGLVITDKA	QSHHETAQSS	TFEGKQVVLQ	AGNDANILGS	NVISDNGTRI
551	QAGNHVRIGT	TQTQSQSEY	HOTQKSLGNS	AGIGFTIGSK	TMTQENQSQS
601	NHTTGSTVGS	LKSDTTIVAS	KHYEQTGSNV	SSPEGNLIS	TSQMDIGAAQ
651	NOLNSKTTGT	VEQKGLTVAT	SSPVTDLAQQ	AIAPAHKAAAN	KSKAKKTAL
701	MPWRLEMQVG	RPKQAKAHK	T*		

This gonococcal protein (ORF115ng-1) shows 91.9% identity with ORF115 over 334aa:

		20	30	40	50	60	70
orf115ng-1.p	NEQTFGEKKVFS	ENGLKLIIN	YWRARRKGHD	ETGHREQNYTL	PFEEITRDIS	LGGSFAYESHK	
65	orf115					STGHSEQNYTL	PREITRNLISLG
							FSFAYESHK
					10	20	30

		80	90	100	110	120	130
5	orf115ng-1.p	ALSRHAPSQGT	ELPQSNRDN	IRTA	KSNGISLPY	TFTPLGSS	SLYIINPANKGYLVET
	orf115	ALSHHAPSQGT	ELPQSN	-----	GISLPYTSNST	ELPSSSLYIIN	FPVNGKGYLVET
		40		50	60	70	80
10	orf115ng-1.p	140	150	160	170	180	190
	orf115	DPRFANYR	OWLGSDYML	GSKLDPNNL	HKRLG	DGYYEQRL	INEQIAELTGHRRLDGYQND
		90	100	110	120	130	140
15	orf115ng-1.p	200	210	220	230	240	250
	orf115	EEQFKALMD	NGATAARSM	NSVGI	ALSAEQAA	QLTSDIVL	VWVQKEVKLPDGGTQT
		150	160	170	180	190	200
20	orf115ng-1.p	260	270	280	290	300	310
	orf115	VYVVRKNG	GIDGKALL	SGSNTQ	INVS	GSLSKNS	SGTIAGRNALINTITL
		210	220	230	240	250	260
25	orf115ng-1.p	320	330	340	350	360	370
	orf115	SAVTATQD	INNIGGIL	SAEQTL	LLNAGNN	INNQS	STAKSSQNAQGSSTYLDRMAGIYT
		270	280	290	300	310	320
30	orf115ng-1.p	380	390	400	410	420	430
	orf115	EKGVLA	QAQGDINI	IAGQIS	NSQSDG	QQTQLQ	AGRDINLDTVQ
35	orf115	EKG	V				

In addition, it shows homology with a secreted *N.meningitidis* protein in the database:

gi|2623258 (AF030941) putative secreted protein [Neisseria meningitidis] Length = 2273

Score = 604 bits (1541), Expect = e-172

Identities = 325/678 (47%), Positives = 449/678 (65%), Gaps = 22/678 (3%)

40	Query: 1	LLVQTEK	DGLHNEQ	TGFEKKV	FSENGK	LHNYWR	ARRK	GHD	ETG	HRE	QNYTL	PEE	ITR	DIS	60
	Sbjct: 739	LIVGT	PE	SAL	DN	DEL	TG	KTI	-TDK	GD	LH	YR	++K	D	TG+
45	Query: 61	LG	SFAY	ESH	SKLS	RHAPS	QGT	ELPQSN	RDN	IRTA	KSNGIS	LPY	TFTPL	GSS	SLYII
	Sbjct: 797	MG	ISAY	KGY	-----	AP	Q	AS	DI	FGT	V	-----	FPV	VA	ENG
50	Query: 121	NP	ANKGYL	VET	DPR	FANYR	OWL	GS	DYML	GSKL	DPNNL	HKRL	G	DGYYEQ	RL
	Sbjct: 841	AP	NNKGYL	IET	DP	FAFT	DYR	KWL	GS	GYML	AAL	QDP	PNH	IHKRL	G
55	Query: 181	GHRRLD	GYQND	EEQ	FKAL	MDNG	ATAAR	SMNL	SVGI	ALSAEQ	AAQLT	SDIVL	VWVQ	KEVKLP	240
	Sbjct: 901	G	YRRLDGY	NDEE	QFKAL	MDNG	T	A	+	L	+	GIAL	SAEQ	A	L
60	Query: 241	DG	GTQTVL	MPQV	YVVR	KN	GGIDG	KALL	SGSNTQ	INVS	GSLSK	N	SGTI	AGR	NAL
	Sbjct: 961	D	GTQTVL	E	YVVR	+	++G	GALL	SGS	I	SG	++N	G	IAGR	ALI
65	Query: 300	LD	NIGRI	HAQ	KS	SAVTATQ	DINNIGGIL	SAEQTL	LLNAGNN	INNQS	STAKSSQ	NAQGS	SSTY	359	
	Sbjct: 1020	+	N	+	G	+	+	A	DI	N	G	I	AE	LLL	A
	Query: 360	LDRMAGI	YITG	KEG	VLA	QAQGD	INI	IAGQIS	NSQSDG	QQTQLQ	AGRD	INL	DTVQ	TKGYK	419
	Sbjct: 1079	I	GRVAGI	YILT	GR	QNGSV	LLD	AGN	IVL	TASE	LT	NQSE	DGQ	TVL	N

Query: 420 EIHFDADNHTIRGSTNEVGSSITQKGDVTLSSGNNLNAAAEVSGAKGTAVYAKNDITI 479
FD+DN+ IR NEVGS+I+T+G++L + ++ +AAEVGS +G L + A DI +
Sbjct: 1139 NTIFDSDNYVIRKEQNEVGSITRTGNLSLNAGDIRIRAAEVGSEQRLKLAAGRIDIV 1198

Query: 480 SSGIHAGQVDDASKHTRSGGGGNKLVITDKAQSHHETAQSSSTFEKGQVVLQAGNDANILG 539
+G +DA K+TGRSGGG K +T ++ + A ST +GK++L +G D + G
Sbjct: 1199 EAGKAHTETEDALKYTRSGGGGIQKQMRHLKNQNGQAVSGTLDGKEIILVSGRDIITVG 1258

Query: 540 SNVISDNGTRIQAQNHVRIGTTQTQSQSEYHQTQKSGLM-SAGIGFTIGSKTNTQENQS 598
SN+I+DN T + A N++ + +T+S+S ++ +KSGLM S GIGFT GK +TQ N+S
Sbjct: 1259 SNIIADNHTILSAKNINVLKAAETRSRAEMNKKEKSGLMSSGGIGFTAGSKKDTQTNRS 1318

Query: 599 QSNHEHTGSTVGS LKGDTTIVASKHYEQTGSNNVSSPEGNNLISTQSMIDIGAAQNQLNSKTT 658
++ HT S VGS L G+T I A KHY QTGS +SSP+G+ IS+ + I AAQN+ + ++
Sbjct: 1319 ETVSHTESVVGSLNGNTLSAGKHYTQTGSTISSPQGDVGISSGKISDAQAQNRYSQESK 1378

Query: 659 QTYEQKGLTVAFSSPVT D 676
Q YEQKG+TVA S PV +
Sbjct: 1379 QVYBQKGVTVVAISVPVNV 1396

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 62

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 517>:

1 ..TCAGGGAATA ACCTCAATGC CAAAGCTGCC GAAGTCAGCA GCGCAAAACGG
51 TACATCTGCT GTGCTCGCCA CACATCAGC GCAGGCATCA
101 ACACGACCCA TCTTGATGAT GCGTCCAAAC ACACAGGCG AGCGGTGCT
151 GGCATAAAT TAGTCATTAC CGATAAAGCC CAAAGTCATC ACGAAGAAGCG
201 CCAAGCAGCG ACCTTTGAAG GCAAGCAGT TGTATTGTCAG GCAGAAACGC
251 ATGCCAACAT CTTTGGCAGC AATGTTATT CCGATAATGG CACCCAGATT
301 CAAGCAGGCA ATCATGTTTG CATTTGTACA ACCCAAACCTC AAGCCAAAGC
351 CGAAACCTAT CATCAAAACC AGAAATCAG ATTGATGAGT CGAGGTATCG
401 GCTTCACTAT TGGCAGCAAG ACAAAACACAC AAGAAAACCA ATCCCAAAAGC
451 AACGAACATA CAGGCAGTAC CGTAGGCAGC TTGAAAGGCG ATACCAACAT
501 TGTTCAGGCG AAACACTACG AACAAATCGG CAGTACCGTT TCCAGCCGCG
551 AAGGCAACAA TACCATCTAT GCCCAAAGCA TAGACATTCA AGCGGCACAC
601 AACCAATTAACACAGTAATAC CACCCAAACC TATGAACAAA AAGG. CTAAAC
651 GTGGGCATTC AGTTGCGCCG TTACCGATTG GGCACACAAA ...

This corresponds to the amino acid sequence <SEQ ID 518; ORF117>:

1 ..SGNNLNAAAEVSSANGTLA VSANNDINIS AGINTTHVDD ASKHTGRSGG
51 GNKLVITDKA QSHHETAQSS TFEKGQVVLQ AGNDANILGS NVISDNGTQI
101 QAQNHVRIGT TQTQSQSEY HQTQKSGLMS AGIGFTIGSK TQNTQENQSQS
151 NEHTGSTVGS LKGDTTIVAG KHYEQIGSTV SSPEGNNIY AQSDIQAHA
201 NKLNSNTTQT YEQKXLTVAF SSPVTDLAQQ ...

Computer analysis of this amino acid sequence gave the following results:

Homology with the pspA putative secreted protein of *N.meningitidis* (accession number AF030941)

ORF117 and pspA protein show 45% aa identity in 224aa overlap:

Orf117: 4 NLNAAAEVSSANGTLAVSANNDINISAGINTTHVDDASKHTRSGGGGNKLVITDKAQSH 63
++ +AAEV S G L ++A DI + AG T +DA K+TGRSGGG K +T ++
pspA: 1173 DIRIRAAEVGSEQRLKLAAGRIDIKVEAGKAHTETEDALKYTRSGGGGIQKQMRHLKNQ 1232

Orf117: 64 HETAQSSSTFEKGQVVLQAGNDANILGSNVISDNGTQIQAGNHVRIGTTQTQSQSEYHQT 123
+ A ST +GK++L +G D + GSN+I+DN T + A N++ + +T+S+S ++
pspA: 1233 NGQAVSGTLDGKEIILVSGRDIITVGSNIIADNHTILSAKNINVLKAAETRSRAEMNK 1292

Orf117: 124 QKSGLM-SAGIGFTIGSKTNTQENQSQSNEHTGSTVGSLKGDTTIVAGKHYEQIGSTVSS 182
+KSGLM S GIGFT GSK +TQ N+S++ HT S VGSL G+T I AGKHY Q GST+S
pnpA: 1293 EKSGLMGSGGIGFTAGSKKDTQNRSETVSHTESVVGSLNGNTLISAGKHYTOTGSTISS 1352

5 Orf117: 183 PEGNNTIYAQSIDIQAAHNKLNSTNTQTVEQKXLTVAFFSSPVTD 226
R+G+ I + I I AA N+ + + Q YEQK +TVA S PV +
pspA: 1353 PQGDVGISGKISIDAAQNRYSQESKOVYEQKGVTVAISVPVVN 1396

Homology with a predicted ORF from *N.gonorrhoeae*

10 ORF117 shows 90% identity over a 230aa overlap with a predicted ORF (ORF117ng) from *N.gonorrhoeae*:

	orf117.pep	SGNNLNKAAAEVSSANGTAVASANDINIS	30
15	orf117.ng	IHFADNHTIRGSTNEVGSSIQTQKGDVTLSSGNNLNKAAAEVGSAGKTLAVYAKNDITIS	480
	orf117.pep	AGINTTHVDVDAKSKHTGRSGGNGKLVITDKAQSHHETAQSSTFEFGQVVLQAGNDANILGS	90
20	orf117.ng	SGIHAGQVDVDAKSKHTGRSGGNGKLVITDKAQSHHETAQSSTFEFGQVVLQAGNDANILGS	540
	orf117.pep	NVISDNGTQIQAGNHVIRIGTQTQTSQSETHYGTQKSGLMSAGIGFTIGSKNTNQENQSQS	150
25	orf117.ng	NVISDNGTRIQAGNHVIRIGTQTQTSQSETHYGTQKSGLMSAGIGFTIGSKNTNQENQSQS	600
	orf117.pep	NEHTGSTVGLSKGDDPTTIVAGHKHYEQIGSTVSSPEGNNNTIYAQSIDQIAAHNKLNSNTTQT	210
30	orf117.ng	NEHTGSTVGLSKGDDPTTIVASKHYEQIGSTVSSPEGNNLISTQSMIDIAAQNLNSKTTTQT	660
	orf117.pep	YEQKXLTIVAFSSPVTDLAQQ	230
30	orf117.ng	YEQKGLTIVAFSSPVTDLAQQAIVAHKAAKQEDKAKTALPMQWRLPMQVGLFKFOAKAPK	720

An ORF117ng nucleotide sequence <SEQ ID 519> was predicted to encode a protein having amino acid sequence <SEQ ID 520>;

	1	..LLVQTEKDIGL	HNBPQTJGEKK	VFSFENGKLLN	YWRARRRGKHD	ETGHEQNYT
35	51	LPBETIRTDLS	LGSRFAYESHSS	KALSRHAPSO	DTQLPQSNRD	NIRTAKEGNY
	101	SLPYTTNSFT	PLPGSSLYII	NPANGKVLGY	TDFPRFANYQ	WLGSDDYMLGS
	151	LKLDPNNLHK	LRGDGYEYQR	LNAQIEALTE	GHRRLDGYNQ	DEEQKALML
	201	NGVTAARSMM	LSGVALLSAE	SQAQILTSVGS	WLVRQKEVLK	DGGTOTVLMF
	251	QVTVRVKNGG	LDGKGLALSG	NNQITNDSDG	LKNSGTIAGR	NALINPTDRI
40	301	DNIGGRTHAQ	KSVAITAQTD	NNIGGIELSA	QTLTLLNAGN	INNQTAKGSS
	351	QNAGGSSTYL	DRMAGIYVTE	KEKGVLAAGA	GKIDNI TAQI	INSQSDGGT
	401	RQAGRDNLIN	DTVTQKYQYE	IHFADNMTHI	RGSTNEVGSN	IQTGKDVTLT
	451	SGNLNNAKAA	EVGSAKATLE	IVHADNITTS	SGIHAGQVDD	ASKHTGRSGG
	501	GKLVITPDKA	QSNQSHTAQS	TDFQGVKVLQ	AGNDANTLGS	NVISDNGTRI
45	551	QAGNIGVQV	TSQDSBETRI	HTQKQSGSLG	AGIGETIGBK	NTIQENGQSS
	601	NHETGSGVGS	IKADPTIVAS	KHEGTQSGNV	SFEGNNLTS	TQSMDDIAAQ
	651	NNLSKKTQTT	YEQKGLIVAF	SPSYTDLQAG	AIAYAWKAS	QDFKAKTAL
	701	MEWRLMVOVG	RELEOKAPAK	K*		

Further work revealed the following gonococcal partial DNA sequence <SEO ID 521>:

50	1	TTGCTTGTGC	AAACCAAGAAA	AGACGGTTTG	CATAACGACG	AAACCTTTGG
	5	CGAAGAGAAA	CTTCTTACGG	ATTTTCGTGA	TTGTGCAGAC	TACTGGGCTG
	101	CGCGTCGTAA	AGGATCATGT	GAACACGGCG	ATCTGTGAACA	AAATTTATCT
	151	TTCTTCGGTA	AAAACACAGT	CGACATCTCA	CGACATCGCT	TTCGCTATAC
55	201	ATCGCTGTGC	ATCGCTGTGC	CGACATCGCT	CGACATCGCT	CGACATCGCT
	251	TGCCAACATG	TAAACGGGGT	ATATCTACCT	TGCAACAAAG	CGACGCTTCT
	301	TGCTTACCTT	TATGCCGCCA	TTCTTTTACC	CCATTACGCT	CGACGACGTT
	351	ATACATTATC	ATCTCTTGCA	ATGAAGGCTTA	TCGTGTTGAA	ACCGATTCCG
60	401	GCTTCTATCA	CATACCTGCA	TTGTTTGGTA	TGTGATCATG	CTGTGGGACG
	451	CTCAACATG	CCACCAATGA	TTCTTTTAAA	TCTTGGGTG	ATGTTATTTA
	501	CGACGACAGT	TTATTCATGT	ACCAATATCA	AGGATCTGCA	GGGATCTGCA
	551	GTTTACAGAG	TTATCAACAG	CGACGAAGAC	AATTTAAGAC	CTTATGSGAT

	601	AATGGCGCGA	CTGGGCGAG	TTCGATGAAT	CTCAGCGTTG	GCATTGCATT
	651	AAGTGGCGAG	CAAGCAGCGC	AACTGACCAG	CGATATTGTT	TGGTGGGTAC
	701	AAAAAGAAGT	TAAACTTCCT	GATGGCGGCA	CACAAACCGT	ATTGATGCCA
5	751	CRGGTTTATG	TACCGGTTAA	AAATGGCGGC	ATGACGGGTA	ARGGTGCATT
	801	GTCTCCAGTC	AGCAATACAT	AAATCAATGT	TTACGCGAGC	CTGAATAACT
	851	CAGGCACAGT	TCCAGGCGGC	AATGGCGTTA	TTATCAATAC	CGATACGGTA
	901	GACAATATCG	GTGGGCGTAT	TCATGCGCAA	AAATCAGCGG	TTACGGCCAC
	951	ACAAGCATCT	AATAATATTG	GCGGCATTCT	TTCTGCGGAA	CAGACATATAT
10	1001	TGCTCAATGC	GGGTAAACAC	ATCAACAACC	AAAGCAGCGC	CAGAGCAGAT
	1051	CAAAATGTCAC	AAGGTAGCAG	CACCTACCTA	GACCGAATGG	CAGGTATTTA
	1101	TATCACAGCG	AAAGAAAAAG	GTGTTTGTAG	AGCGCAGGCA	GGCAAAAGCA
	1151	TCAACATCAT	TGCGGGTCAA	ATCAGCAATC	AATCAGATCA	AGGGCAAAAC
	1201	CGGCTGCAGG	CAGGACGCGA	CATTAACTGT	GATACGGTAC	AAACCGGCCA
15	1251	ATATCAAGAA	ATCCATTTTG	ATGCCGATAA	CCATACCATC	CGAGGTTCAA
	1301	CGACGGAAGT	CGGCAGCAGC	ATTCAAACAA	AAGGCGATGT	TACCCtatTG
	1351	TCAGGGAATA	ATCTCAATGC	CAAAGCTGCC	GAAGTCGGCA	GGGCAAAAGG
	1401	CACACTTGCC	GIGTATGCTA	AAAATGACAT	TACTATCAGC	TCAGGCTATCC
	1451	ATGCGGCGCA	AGTTGATGAT	CGGTCCAAAC	ATACAGGCGG	AAGCGGCGGC
20	1501	GCTAATAAAT	TAGTCTATAC	CGATAAAGCC	CAAGTCTATC	ACGAACATGC
	1551	TCRAAGCAGC	ACCTTTGAAG	CAACGAAAGT	ACTATTGCGA	GCAGAAACGC
	1601	ATGCCAAACAT	CCTTGGCAGT	AATGTTAATT	CCGATAATGG	CACCCGGAAT
	1651	CAAGCAGGCA	ATCATGTTGC	CATTGGTACA	ACCCAACTC	AAAGCCAAAG
25	1701	CGAAACCTAT	CATCAAAACC	AAAATCAGG	ATTGATGAGT	GCAGGTATCG
	1751	GCTTCACTAT	TGGCAGCAG	ACAAACACAC	AAGAAACCA	ATCCCAAAGC
	1801	AACGAACATA	CAGGCAGTAC	CGTAGGCAGC	CTGAAAGCGC	ATACCAACAT
	1851	TGTTGCAAGC	AAACACTACG	ACAAACCGG	CAGCAAGTTC	TTCAGCGCCG
	1901	AGGGCAACAA	CCTTATCAGC	ACGCAAAAGT	TGGATATTGG	CGCAGCACAA
	1951	AACCAATTAA	ACAGCAAAAC	CACCCAAACC	TACGAACAAA	AAGGCTTAAC
30	2001	GGTGGCATTG	AGTTGCGCCG	TTACCGATTG	GGCAACACAA	CGGATTTCCG
	2051	TAGCACACAA	AGCAGCAAAC	AAGTCGGACA	AAGCAAAAAC	CACGCGGTTA
	2101	ATSCCATGGC	GGCTGCCAAT	GCAGGTTGGC	AGGCCTATCA	AACAGGCCAA
	2151	GGCGCACAAA	ACTTTG			

This corresponds to the amino acid sequence <SEQ ID 522; ORF117ng-1>:

	1	LLVQTEKDGL	HNEQTFGKK	VFSENGKLHN	YWRARRKGD	ETGHRQNYT
35	1	LPEEITRDIS	LGSFAYESH	KALSRHAPSQ	GTELPCSNRD	NIRTAKSNGI
	101	SLPYTFNSFT	PLPGSSLYII	NPANKGYLVE	TDFRFANYRG	WLGS DYMLGS
	151	LKLPDNNLHK	RLGGDYVEQR	LINEQIAELT	GHRRLDGYGN	DEEQFKALMD
	201	NGATAARSMN	LSVGIALSAE	QAQLTSDIV	WLVOKEVKLP	DGGTQTVLMP
40	251	QVYVRVKNNG	IDKGALLSG	NTQINVSGS	LKNSGTIAGR	NALIINTDTL
	301	DNIGGRIHAQ	KSATVATQDI	NNIGILLSAE	QTLINLAGNN	INNQSTAKSS
	351	QNAQGSSTYL	DRMAGIYITG	KEKGVLAQA	GKDINIAGQ	ISNQSDQGGT
	401	RLQAGRDINL	DTVQTKYQE	IHFADNHTI	RGSTNEVGS	ISQTKGDTVLL
	451	SGNNIMAKAA	EVSGAKGTLA	YVAKNDITIS	SGTHAGQVDD	ASKHTGRSGG
45	501	GNKLVTIDTA	SGSHETAGSS	TFGKQVVLG	AGNDANILGS	NVISDNQTRI
	551	QAGNHVRITG	TQTQSOSETY	HPQKSGKLMS	AGIGFTIGSK	NTQENQSQS
	601	NEHTGSITGS	LKGDITIVAS	KHYETQGSNV	SSPEGNLLIS	TQSDMDIGAAQ
	651	NQLNSKTTQT	YEQKGLTVAF	SSPVTDLAQG	AIIVAHKAAN	KSKDAKATTL
	701	MPWRLEMGVG	RPIKQAKAHK	T*		

ORF117ng-1 shows the same 90% identity over a 230aa overlap with ORF117. In addition, it shows homology with a secreted *N.meningitidis* protein in the database:

gi|2623258 (AF030941) putative secreted protein [Neisseria meningitidis]Length = 2273

Score = 604 bits (1541), Expect = e-172

Identities = 325/678 (47%), Positives = 449/678 (65%), Gaps = 22/678 (3%)

Query: 1 LLVQTEKDGLHNEQTFGKKVFSENGKLHNYWRARRKGDHETGHRQNYTLPEEITRDIS 60
L+V T+ L N+T+T G K + ++ G L H Y R +KG D TG+ Y E++ I

Sbjct: 739 LLVGTPEESALDNDTETLGTKTT-TDKGDLHRYHRHKKGRDSTGYRSPEYAPAEVS-SIR 796

Query: 61 LGSFAYESHKALSRHAPSQGTLEPQSNRDNIRTAKSNGISLPHYTFNSFTPLPGSSLYII 120
+G AY+ + AP Q ++P + + NGI +T LP SSL+ I

Sbjct: 797 MGISAYKG-----AFQQAASIDFGT---VFPVAENGHPTPT-----LNSSLFAT 840

Query: 121 NPANKGYLVETDFRFANYRQWLGS DYMLGSLKLPDNNLHKRLGGDYVEQRLINEQIAELT 180

P NKGYL+ETDP F +YR+WIGS YML +L+ DPN++HKRLGDGYEQL+NEQIA+IT
 Sbjct: 841 APNNKGYLIETDPAFTDYRKNLGSGVYMLAALQQDPNHHKRLGDGYEQLVNEQIAKIT 900
 Query: 181 GHRRLDGYQNDDEEQFKALMDNGATARSMLNSVGLASAEQAQQLTSDIVWLQKEVKLP 240
 5 +*RRRLDSY NDEEQFKALMDNG T A+ + L+ GIALSSEQ R+LTSIDIWL + V LP
 Sbjct: 901 GYRRLDGTNDDEEQFKALMDNGITIAKELQTLFGIALSSEQVARLTSIDIWLENETVTLP 960
 Query: 241 DGGTQTVLMPQVYVVRVKNVGGIDGKALLSGSNTQINVSGLKN-SGTIAGRNALINTDT 299
 10 DG TQTVL P+VYVR + ++G+GALLSGS I SG++N G IAGR ALI+
 Sbjct: 961 DGGTQTVLKPKYVVRARPKDMNGQGALLSGSVVDIG-SGAIENRGGLIAGREALILNAQN 1019
 Query: 300 LDNIGGRIHAQKSAVTATQDINNIGILSAEQTLNLAGNNINNNQSTAKSSQNAQGSSTY 359
 + N+ G + + A DI N G I AE LLL A NNI ++S +S+QN QGS
 15 Sbjct: 1020 IKNLQGD LGQKNIFAAAGSDITNTGSI-GAENALLKASNNIESRSETRSNQNEQGSVRN 1078
 Query: 360 LD RMAGIYITGKEKGVLAQAQGDINI IAGQISNQSDQGGTQLRQAGRDINLOTVTGQKYQ 419
 + R+AGIY+TG++ G + AG +I + A +++NQS+ GQT L AG DI DT + Q
 Sbjct: 1079 IGRVAGIYLTGRQNGSVLLDAGNNIVLTASELTNQSEGDGQTVLNAAGGDIRSDTTGISRNQ 1138
 Query: 420 EIHFADNHTIRGSTNEVGSSTQTKGDVTLTLLGNNLNAKAAEVGSAKGTFLAVYAKNDITI 479
 20 FD+DN+ IR NEVGS+I+T+G++L + ++ +AAEVGS +G L + A DI +
 Sbjct: 1139 NTIFDSNYVIRKEQNEVGSSTIRTRGNLSLNKAGDIRIRAAEVGSEQGRKLKLAAGRDIVK 1198
 Query: 480 SSGIHAGQVDDASKHTGRSGGGNKLVITDKAQSHHETTAQSSTFEKGQVVLQAGNDANILG 539
 25 +G + +DA K+TGRSGGG K +T ++ + A S T +GK++L +G D + G
 Sbjct: 1199 EAGKAITEDEDALKYTRSGGGIKQKMTIRHLKNQNGQVSGTLDGKEILVSGRDITVTG 1258
 Query: 540 SNVISDNQTRIQAGNHVRIGTTQTQSQSETYHOTQKSGML-SAGIGFTIGSKTNTQENQS 598
 30 SN+I+DN T + A N++ + +T+S+S ++ +KSGML S GIGFT GSK +TQ N+S
 Sbjct: 1259 SNIIDNHTILSAKNNIVLKAATERSRSAEMNKKEKSGMLSGGGTIGFTAGSKKTDQTNRS 1318
 Query: 599 QSNHGTSTVGS�KGDITIVASKHYEQTGSNVSSPEGNNLISTQSMIDGAAQNNQINSKTT 658
 35 ++ HT S VGS L G+T I A KHY QTGS +SSP+G+ IS+ + I AAGN+ + ++
 Sbjct: 1319 ETVSTSVVGSINGNTLISAGKHYYTQTGSTISSFGQDVGISSGKISIDAAQNRYSQESK 1378
 Query: 659 QTYEOKGLTVAFSSPVTD 676
 Q YEOKG+TVA S FV +
 Sbjct: 1379 QVYEOKGVTVAISVFPVN 1396

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 63

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 523>:

1 ATGATTTCAC TCGTACTGTT TCTAGCTGTC GTCCTCGCGG TTGTGCGCTA
 51 CAACATGTAT CAGGAAAACC AATACCGCAA AAAAGTGGCG GACCGATTGC
 45 101 GACACTCCGA CAAAGATGCC CTGCTCAACA GCWAAACCAG CCAATGTCGCG
 151 GACGCGAAAC CGTCCGCGCG GTCACTCATG ATGCGCGAAAC CCCACCGCG
 201 GGTCAAAAAC ACGGCAAAAC CCCAAGACCC CGYCATGCCC AACCTGCAAG
 251 AACAGGATGC CGTCTACATC GCCAAGCAGA AACAGGCAAA AGCCTCCCGC
 301 TTCAAAACCG AAATCGAAAC CGCCTTGGAA GAAAGCGGCA TTATCGGCAA
 50 351 CTCGCGCCAC ACCTGTTTCCG AACCCCAAAC CGGACATTCC GCAACGAAAC
 401 CTGCGCAGCG GTGCGCAAAA CTTGCAACCG TTCCGCAAAC ACCTCGAAAC
 451 CCGCTGATTA CGCTCAAGAA ACTGTCAAAA GTCGAATTAT CCTGGTTTGA
 501 CTGCGCAGCG GACTTCATCT CCTAT...

This corresponds to the amino acid sequence <SEQ ID 524; ORF119>:

1 MIYIVLEFLV VLAVVAYNMY QENQYRKKVR DQFGHSDKDA LLNSXTSHVR
 51 DGKPSGGSVN MPKQPAVKR TAPKODEXMR NLQEQDAVYI AKQKQAKASP
 101 FKTEIETALE ESLIGNSAH TVSEPQTGHS ATPADASAK FAPVQTPAK
 151 FLITLKLKSL VELSWFDVRI DFISY...

Further work revealed the complete nucleotide sequence <SEQ ID 525>:

1 ATGATTTACA TCGTACTGTT TCTAGCTGTC GTCTCTGCGG TTGTGCGCTA
 51 CAACATGTAT CAGGAAAACC AATACCGCAA AAAAGTGGCG GACCAAGTTCG
 101 GACACTCCGA CAAAGATGCC CTGCTCAACA GC AAAACAGG CCATGTCCGCG
 151 GACGGCAAAAC CGTCCGGGCG GTACGTCATG ATGCCGAAAC CCCAACCGCG
 201 GGTCAAAAAA ACGGCAAAAC CCCAAGACCC CGCCATGCGC AACCTCGAAG
 251 AACGAGATCG CGCTGATCAT CCGTCAAGCA AACCTCGGAA AGCCTCCGCG
 301 TTCAAAACCG AAATCGAAAC CGCCTTGGAA GAAAGCGGCA TTATCGGCA
 351 CTCCGCCACC ACGGTTTCCG AACCCCAAC CGGCAATTCC GCACCGAAAC
 401 CTGCGGACGC GCGGCAAAAC CCTGCACCG GTCCGAAAC ACCTGCABAA
 451 CGCTGATTA GCTCAAAAGA ACTGTCAAAA GTCCGAATTAC CCGTGGTTTGA
 501 CGTGGCTTC GACTTCATCT CCTATATCGC GTGACCGGAA GCCAAAGAAC
 551 TGCAGCGACT GCGGCGCCTT TCCAACGCT GCGGCTAACA GATTGTGCGC
 601 TGCACCATGG ACGACCATTT CAGATTGCGC GAACCCATCC CGGGCATCCG
 651 CTATCAGGCA TTATCGTGG GTATTGAGCG AGTCAGCGCG AACGGAATTTG
 701 CCTCGCAGGA AGAACTCTCC GCATTCAACC GCCAGGTGGA CGCATTTCGA
 751 CAAGCATGG GCGGTGAGC GCTGCACACC GACCTTGCGC CCTTTATCGA
 801 AGTGCGTTC GCACTGGAGC CATTTCTGCG GCGCGTCGAC CAGACCATCG
 851 CCATCCATT GGTTCGCCCG ACCAGCATCA GCGCGGTAGA ACTGGGTTC
 901 GCCGTAAACG CGGTGGGTTT GCTTTTGGAA GACGACGCGG GTTCCACTA
 951 TACCGACAG TCGGGCTCGA CCGTGTCTC CATCTGCTCG CTCACACAGC
 1001 AGCGGTTTAC CAACGCGCTT TTGACACCC AGTCCTACAA AGCCTTCAGT
 1051 ATGCTGCTCG ACATCCCGCA CTCTCGGCA GCGGAAAAA CTTTGAAGCA
 1101 TTTGTTTATG GATTGCGCG TACGCTGTC CGGCCAGTTG AACCTGAATC
 1151 TGGTCAACGA CAAATGGAA GAAGTTTGA CCAATGGCT CAAAGAGGTG
 1201 GCACTTATG TATTGGCGG TCAGTCCGAG ATGCTCAAG TCGGTATCGA
 1251 ACGGGGCGG AAAACCGCAT TGCGCTGTT CTCTCAA

This corresponds to the amino acid sequence <SEQ ID 526; ORF119-1>:

1 MIYIVFLAV VLAVVAYNMY QENQYRKVR DQFGHSDKDA LLNSKTSHVR
 51 DGKPSGGGVM MPKPQPAVKK TAKFQDPAMR NLQEQDAVYI AKQKQAKSP
 101 FKTEIETALE ESGIIGNSAH TVSEPTGHS APKPADAPAK PAPVPQTPAK
 151 PLITLKLKSK VELFWDFVR DFISYIALTE AKELHALPRL SNRCRYQIVG
 201 CTMDHFQIA EPIPIGIRYA FIVGIQAVSR NGLASQELLS AFNRQVDAFA
 251 QSMGCGTHT DIAAFISVAS ALDGCARDQ TTAHLVSP TSISGVELRS
 301 AVTGVGFVLE DGRFPHYDT SGTSMESTCS LNEEPTNL LDNQSYKGS
 351 MLLDIPHSPA GEKTFDDLPM DLAVRLSGQL NLNLVNDKME EVSTQWLKDV
 401 RTYVLARQSE MLKVGIEPBG KTALELRLFS*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF119 shows 93.7% identity over a 175aa overlap with an ORF (ORF119a) from strain A of *N.*

meningitidis:

		10	20	30	40	50	60
orf119.pep	MIYIVFLAVVLAVVAYNMY	QENQYRKVRDQFGHSDKDALNSXTS	SHVRD	GKPSGGGVM			
orf119a	MIYIVFLAAVLAVVAYNMY	QENQYRKVRDQFGHSDKDALNSKTS	SHVRD	GKPSGGGVM			
		10	20	30	40	50	60
		70	80	90	100	110	120
orf119.pep	MPKPQPAVKKTAKPQDP	PMRNLOEQDAVYIAKQKQAKSP	FKTEIETALEESGIIGNSAH				
orf119a	MPKPQPAVKKTAKSQDP	AMRNLOEQDAVYIAKQKQAKSP	FKTEIETALEESGIIGNSAH				
		70	80	90	100	110	120
		130	140	150	160	170	
orf119.pep	TVSEPTGHSATKPADASAK	PAPVPQTPAKPLITLKLKSK	VELSWDFVRIDFISY				
orf119a	TVPEPTGHSAPKPADAPAK	PVPQTPAKPLITLKLKSK	VELFWDFVRDFISYIALTE				
		130	140	150	160	170	180
orf119a	AKELHALPRLSNRCRYQ	IVGCTMDHFQIAEPIPIGIRYA	FIVGIQAVSRNGLASQELLS				
		190	200	210	220	230	240

The complete length ORF119a nucleotide sequence <SEQ ID 527> is:

1 ATGATTTACA TCGTACTGTT CCTCGCCGCC GTCTCTGCCG TTGTGCGCTA
 51 CAATATGTAT CAGGAAAACC AATACCGCAA AAGAGTGGCG GACCACTTCG
 101 GGCACCTCGA CAAGAATGCC CTGCTCAACA GCAAAACGAG CCATGTCCGC
 151 GACGCGAAAC CGTCCGGCGG CCGAGTCTAG ATGCCGGAAC CCAACCGCGC
 201 GGTCAAAAAC ACGGCAAAAT CCGAAGACCG GCGCATGGCG ARCCCTCGAG
 251 AGCAGGATGC CGTCTACATC GCCAAGCAGA AACAGGCAAA AGCCTCCCCG
 301 TTCAAAACCG AAATCGAAAC CGCCTTGAA GAAAGCGCA TTATCGGCAA
 351 CTCGCCCCAC ACCGTTCCCG AACCCCAAAC CGGACATTCG GCACCAAAAC
 401 CTGCGCAGCG CGCGGCAAAA CCGTTCCCG TTCCGCAAC GCGGCGAAAA
 451 CCGCTGATTA CGCTCAAGA GCTGTGAGAG GTGAGCTCGC CCGTGGTTGA
 501 CGTGGCGTTC GACTTCATCT CTTATATCGC TTCCGACGAA GCCAAGAAC
 551 TGCACGCACT GCGCGCGCTT TCCAACCGCT GCGGCTACCA GATTGTGCGG
 601 TGCACCATGG ACGACCATTT CCAGATTGCC GAACCATTCG CGGGCATCCG
 651 CTATCAGGCA TTTATCGTGG GTATTGAGCG AGTCAGCGCG AACGGACTTG
 701 CCTCGCAGGA AGAAGCTTCC GCATTCAACC GCCAGTGGG TGCATTTCGA
 751 CACAGCATGG GCGGTGAGAC GGTGCACACC GACCTTGCGG CTTTATGACA
 801 AGTGCGCTTC GCACTGGAGC CATTCTGGCG GCGGTGAGC CAGACTATCG
 851 CCATCCATTT GGTTCGCCCG ACCAGCATCA GCGGCGTAGA ACTGCGTTCG
 901 GCGCTAACGG CGTSGSGTTC CGTTTTGAA GACGAGCGCG GGTTCACCTA
 951 TACGACACAG TCGGCTCGGA CACGTTCTTC CATCTGCTCG CTCACACACG
 1001 AGCGGTTTAC CAATGCCCTT TTGACACACC AGTCCTATAA AGGCTTCAGT
 1051 ATGCTGCTCG ACATCCCGCA CTCTCCGCGA GCGCAAAAAA CCTTCGAGCA
 1101 TTTGTTTATG GATTGGCGGG TACGCTGTCT CGGCCAGTTG AACCTGAATC
 1151 TGGTCAACGA CAAAATGGAA GAAGTTTCGA CCAATGGCT CAAAGACGTG
 1201 CGCACTTATG TATTGGCTCG TCAGTCCGAG ATGCTCAAG CGGSTATCGA
 1251 ACGGGGCGCG AAAACGCGAT TGGCGCTGTT CTCCTAA

This encodes a protein having amino acid sequence <SEQ ID 528>:

1 MIYIVLFLAA VLAVVAYNMY QENQYRKQVR DQFGHSDKDA LLNSKTSHVR
 51 DGKFSGGPVM MPKQFAVKK TAKSQDPAMR NLQEQDAVYI AKQKQAKSP
 101 FKTEIETALE ESGLIIGNSAH TVPEPQTHGS APKPADAPAK PVVPQTPAK
 151 PLITLKLKSL VELEWDFVRF DFISYIALTE AKELHALPRL SNRCRYQIVF
 201 CTMDDHFIQA EPIPGIRYQA FIVGIQAVSR NGLASQELS APNRQVDFA
 251 HSMGGTTLHT DLAFIEVAS ALDAFCARVD QTAIHLVSP TSSISGVELRS
 301 AVTGVGFVLE DDGAFHYDT SGTSMESICS LNNPPTTVAL LDNQSYKGSF
 351 MLLDIHPSA GKTFHDLFL DLAVRLSQSL NLNLVNDKME EVSTQWLKDV
 401 RTYVLARQSE MLKVIEEFG KTAIRLFS*

ORF119a and ORF119-1 show 98.6% identity in 428 aa overlap:

		10	20	30	40	50	60
40	orf119a.pep	MIYIVLFLAAVLAVVAYNMYQENQYRKQVRDQFGHSDKDAL	LLNSKTSHVRD	GKPSGGPVM			
	orf119-1	MIYIVLFLAAVLAVVAYNMYQENQYRKQVRDQFGHSDKDAL	LLNSKTSHVRD	GKPSGGPVM			
		10	20	30	40	50	60
45	orf119a.pep	MPKQFAVKKTA	KSQDPAMRN	LQEQDAVIA	KQKQAKSP	FKTEIETALE	ESGIIGNSAH
	orf119-1	MPKQFAVKKTA	KSQDPAMRN	LQEQDAVIA	KQKQAKSP	FKTEIETALE	ESGIIGNSAH
		70	80	90	100	110	120
50	orf119a.pep	TVPEPQTHGSA	PKPADAPAK	PVVPQT	PAKPLITLKLK	SKVELWDF	VRFDFISYIALTE
	orf119-1	TVPEPQTHGSA	PKPADAPAK	PVVPQT	PAKPLITLKLK	SKVELWDF	VRFDFISYIALTE
		130	140	150	160	170	180
55	orf119a.pep	AKELHALPRL	SNRCRYQIV	GCTMDDHFI	QAEPIPGIRY	QAFIVGIQAV	SRNGLASQELS
	orf119-1	AKELHALPRL	SNRCRYQIV	GCTMDDHFI	QAEPIPGIRY	QAFIVGIQAV	SRNGLASQELS
		190	200	210	220	230	240
60	orf119a.pep	AFNRQVDFA	HSMSGGQTL	HTDLAA	FTIEVASAL	DAFCARVDQ	TAIHLVSP
	orf119-1	AFNRQVDFA	HSMSGGQTL	HTDLAA	FTIEVASAL	DAFCARVDQ	TAIHLVSP
		250	260	270	280	290	300
65	orf119a.pep	AFNRQVDFA	HSMSGGQTL	HTDLAA	FTIEVASAL	DAFCARVDQ	TAIHLVSP
	orf119-1	AFNRQVDFA	HSMSGGQTL	HTDLAA	FTIEVASAL	DAFCARVDQ	TAIHLVSP
		250	260	270	280	290	300

		310	320	330	340	350	360
5	orf119a.pep	AVTGVGVLEDDGAFHYTDTSGSTMFSICSLNNEPFTNALLDNQSYKGFSMLLDIHPSPA					
	orf119-1	AVTGVGVLEDDGAFHYTDTSGSTMFSICSLNNEPFTNALLDNQSYKGFSMLLDIHPSPA					
		310	320	330	340	350	360
10	orf119a.pep	GEKTFDDL FMDLAVRLSGQLNLNLVNDKMEEVSTQWLKDVRTYVLARQSEMLKVGIEPPG					
	orf119-1	GEKTFDDL FMDLAVRLSGQLNLNLVNDKMEEVSTQWLKDVRTYVLARQSEMLKVGIEPPG					
		370	380	390	400	410	420
15	orf119a.pep	429					
	orf119-1	KTALRLFSX					

Homology with a predicted ORF from *N.gonorrhoeae*

20 ORF119 shows 93.1% identity over a 175aa overlap with a predicted ORF (ORF119ng) from *N.gonorrhoeae*:

	orf119.pep	MIIVFLAVVLAVVAYNMVQENQYRKVRDQFGHSDKDALLSXTSVHRDGPSSGGSVM	60
	orf119ng	MIIVFLAVVLAVVAYNMVQENQYRKVRDQFGHSDKDALLSXTSVHRDGPSSGGSVM	60
25	orf119.pep	MPKPQPAVKKTKAKPQDPKMRNLQEADAVYIAKQKQAKASPFKTEIETALEESGIIGNSAH	120
	orf119ng	MPKPQPAVKKPAKPDQSAMRNLQEADAVYIAKQKQAKASPFKTEIETALEEIGIIGNSAH	120
30	orf119.pep	TVSEPTGHSATKPADASAKPAEVPQTAKPLITLTKELSKVELSWFVDRIDFISY	175
	orf119ng	TVSEPTGHSAPKPADAPKEVEVPQTAKPLITLTKELSKVELSWFVDRIDFISYALTE	180

The complete length ORF119ng nucleotide sequence <SEQ ID 529> is:

35	1	ATGATTAC	TGCTACTGTT	CCTCGCGCGC	GTCTCGCGCG	TTGTGCGCTA
	51	CAATATGTAT	CAGGAAAACC	AATACCGCAA	AAAAGTGGCG	GACAGTTTCG
	101	GACACTCGGA	CAAGATGCCC	CTGCTCAACA	GCAAAAACAG	CCATGTCCGC
	151	GACGCGAAAC	CGTCCGCGCG	GCCAGTCACT	ATGCGGAAAC	CCCAACCGGC
	201	GGTCAAAAAA	CGGCGCAAAC	CCCAAGACTC	CGCATGCGC	AACCTGCAG
40	251	AACAGGATGC	CGTCTACATC	CCAAGCAGA	AACAGGACAA	AGCCTCCCGC
	301	TTCAAAACGC	AAATCGAACC	CGCCTTGAG	AAAATCGGCA	TTATCGGCAA
	351	CTCGCGCAC	ACGCTTTCG	AACCCCAAC	CGGACATTC	CCACGAAAC
	401	CTGCGCAGC	CGCGGCAAAA	CCGCTTCGCG	TTGCGCAAC	GCGCGAAA
	451	CGGCTGATTA	CGCTCAAAAG	GCTGTGGAAG	GTCGAGCTGC	CCTGCTTGA
45	501	CGTGGCGTtc	gaCTTCATCT	CTATATGCG	GCTGACCGAA	GCCAAAGAAC
	551	TGCAAGCACT	GCGCGCGCTT	tccAAACGCT	GCGGCTACCA	GATTGTGCGC
	601	TGCACCATGG	ACGACCAATT	CCAGATTGCG	GAACCCATCC	CGGGCATCOG
	651	CTATCAGGCA	TTTATCGTGG	GTATCCAGGC	AGTCAGCGCG	AACGACTTGC
	701	CCTCGCAGGA	AGAACTCTCC	GATTTCAACC	GCGAGCGGGA	CGCATTTCGA
50	751	CAAAGCATGG	GCGGTGAGAC	GCTGCACACC	GACCTTTCGCG	CCTTTATCGA
	801	AGTGGCTTCC	GCACTGGAAG	CATTCTGCGC	GCGGCTGAC	CAGACCATOG
	851	CCATCCATTT	GGTTTCGCGC	ACCAGCATCA	GCGGCTGAG	ACTGCGTTCC
	901	GCGGTACACG	GCGTGGGTTT	CGTTTGGGAA	GAGCAGCGCG	CGTTCCACTA
	951	TACGCAACGC	TGSGGCTCGA	CCATGTTCTC	CATCTGCTCG	CTCAACAACG
55	1001	AGCGGTTTAC	CAGTGCCTT	TTGGACAACC	AGTCTACATA	AGGCTTCAGT
	1051	ATGCTGCTCG	ACATCCCGCA	CTTCCGCGCA	GCGGCAAAAA	CCTTCGACGA
	1101	TTTGTPTTATG	GATTTCGCGG	TAGCCTGTCT	CGGTCAAGTG	AACTGTAATC
	1151	TGGTCAACGA	CAAAATGSGAA	GAAGTTTTCGA	CCAAATGGCT	CAAAGACGTA
	1201	CGCACTTATG	TATTGGCGCG	TCAGTCCGAG	ATGCTCAAGT	TCGGTATCGA
	1251	ACCGGCGCGC	AAAACCGGCC	TGCGCTGTTT	TTCATAA	

60 This encodes a protein having amino acid sequence <SEQ ID 530>:

1	MIIVFLAA	VLA VVAYNM	QENQYRKVR	DQFGHSDKA	LLNSKTSVHR
51	DGKPSGGPVM	MPKPQPAVKK	PAKPDQSAMR	NLGEQDAVYI	AKTKQAKASP

5

101	FKTEIETALE	EIGIIGNSAH	TVSEPTGHS	APKPADAPAK	PVPVPQTPAK
151	PLITLKELSK	VELFWFDVRF	DFISYIALTE	AKELHALPRL	SNRCRYQIVG
201	CTMDHDFQIA	EPFPGIRYQA	FIVGIQAVSR	NGLASQEELS	AFNRQADAF
251	QSMGGQT LHT	DLAAFIEVAS	ALDAFCARVD	QTAIHLVSP	TSISGVELRS
301	AVTGVGVLE	DDGAFHYTDT	SGSTMFISCS	LNNEPFTNAL	LDNQS YKGS
351	MLLDIPHSPA	GEKTFDDLFM	DLAVRLSGQL	NLNLVNDKME	EVSTQWLKDV
401	RTYVLARQSE	MLKVGIEFGG	KTALRLFS*		

ORF119ng and ORF119-1 show 98.4% identity over 428 aa overlap:

		10	20	30	40	50	60
10	orf119ng	MIYIVLFLAVLAVVAYNMYQENQYRKVRDQFGHSDKDALLNSKTS	SHVRDQKPGSGGVM				
	orf119-1	MIYIVLFLAVLAVVAYNMYQENQYRKVRDQFGHSDKDALLNSKTS	SHVRDQKPGSGGVM				
		10	20	30	40	50	60
15	orf119ng	MPKPQPAVKKPAKPQDSAMRNLOEQDAVYIAKQKQAKASPFKTEIETALEE	EIGIIGNSAH				
	orf119-1	MPKPQPAVKKPAKPQDPAMRNLOEQDAVYIAKQKQAKASPFKTEIETALEE	SGIIGNSAH				
		70	80	90	100	110	120
20	orf119ng	TVSEPTGHSAPKPADAPAKPVPVPQTPAKPLITLKL	SKVELPWFDFVDFDISYIALTE				
	orf119-1	TVSEPTGHSAPKPADAPAKPVPVPQTPAKPLITLKL	SKVELPWFDFVDFDISYIALTE				
		130	140	150	160	170	180
25	orf119ng	AKELHALPRLSNRCRYQIVGCTMDHDFQIAEPFPGIRYQAFIVG	IQAVSRNGLASQEELS				
	orf119-1	AKELHALPRLSNRCRYQIVGCTMDHDFQIAEPFPGIRYQAFIVG	IQAVSRNGLASQEELS				
		190	200	210	220	230	240
30	orf119ng	AFNRQADAFQSMGGQTLHTDLAAFIEVASALDAFCARVQDTIAIHLV	SPTSISGVELRS				
	orf119-1	AFNRQADAFQSMGGQTLHTDLAAFIEVASALDAFCARVQDTIAIHLV	SPTSISGVELRS				
		250	260	270	280	290	300
35	orf119ng	AVTGVGVLEDDGAFHYTDTSGSTMFISCSLNNEPFTNALLDNQ	SYKGSMLLDIPHSPA				
	orf119-1	AVTGVGVLEDDGAFHYTDTSGSTMFISCSLNNEPFTNALLDNQ	SYKGSMLLDIPHSPA				
		310	320	330	340	350	360
40	orf119ng	GEKTFDDLFMDLAVRLSGQLNLNLVNDKMEEVSTQWLKDVRTYVLARQ	SEMLKVGIEFGG				
	orf119-1	GEKTFDDLFMDLAVRLSGQLNLNLVNDKMEEVSTQWLKDVRTYVLARQ	SEMLKVGIEFGG				
		370	380	390	400	410	420
45	orf119ng	KTALRLFSX					
	orf119-1	KTALRLFSX					
		429					
50	orf119ng	KTALRLFSX					
	orf119-1	KTALRLFSX					

55 Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 64

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 531>

1 ..GGCGGCACG GCACGGAAGA TTCTTTCATG AACACACGCG ACAC.ATCAG
 51 GCAGATAGTC GAAAGCACCA CCGGTACGAT GAAGCTGCTG ATTTCTCCCA
 101 TCGCCTTGAT TTCATTGGTA GTCCGGCGCA TCGCGTGAT GAACATCATG
 151 CTGGTGTCCG TTACCGAGCG CACCAAGAA ATCGGCATAC GGAATGGCAAT
 201 CGCGCGCGCG CCGCGCAATA TTTCGACGA GTTTTGTGAT GAGCGGTGT
 251 TATCTGCGT CPTGCGCGG TGGGTTTGTC GCGCGCGGTC CCGCGCGGTC
 301 AGCTCTCGT TCATTCATT TGTAAACGAC TTCGCGGTG ACATTTCGCG
 351 CATGTCCGTC ATCGCGCGCG TCGCTGTTC GACCGGAATC GGCATCGCT
 401 TCGGCTTTAT GCGTGCCAAT AAAGCAGCCA AACTCAATCC GATGACGCA
 110 451 TTGGCACAGG ATTGA

This corresponds to the amino acid sequence <SEQ ID 532; ORF134>:

1 ..ARHGTEDEFFM NNSDXIRQIV ESTTGTMKLL ISSIALISLV VGGIGVMNIM
 51 LVSVTERKE IGIRMAIGAR RGNIXQFLI EAVLICVIGG LVGVLSAAV
 101 SLVNFHFTD FPMDISAMSV IGAVACSTGI GIAFGMPAN KAAKLNPIDA
 151 LAQD*

Further work revealed the complete nucleotide sequence <SEQ ID 533>:

1 ATGTCGGTGC AAGCAGTATT GGCGCACAAA ATCGTTCGCG TTCTGACGAT
 51 GCTCGGCATC ATCATGGTA TCGCGTCGGT GGTTTCGCTG GTCGCAATTGG
 101 GCAATGGTTC CGAGAAAAAA ATCCTTGAAG ACATCAGTTC GATAGGGACG
 151 AACACCATCA GCATCTTCCC GGGGCGCGCG TTGCGGACAA GCGCGAGCGG
 201 CAGGATTAAC ACCCTGACCA TAGACGACGC AAAAATCATC GCCAACAATA
 251 GCTAAGTTGC TTCGCGCACG CCGATCTT CGACCGCGCG CAGCTGACT
 301 TACCGACAAC CCGACCTGAC CGCTCGCTT TACGCGGTG CGGACAATA
 351 TTTCGACGTG CGCGGACTGA AGCTGGAAAC GGGGCGGCTG TTTCGACAAA
 401 ACGATGTGAA AGAAGACGCG CAGGTCGTGC TCATCGACCA AAATGTCAAA
 451 GACAAACTCT TTGCGGACTC GGAATCGTTG GGTAAACACA TTTTGTTCAG
 501 GAAACGCCCC TTGACCGTCA TCGCGGTGAT GAAAAAGAC GAAACGCTT
 551 TCGGCAATTG CGACGTGCTG ATGCTTTGGT CGCCTATAC GACGGTGATG
 601 CACCAATACA CAGCGAGAG CACACCAAC TCATCACCG TCANAATCAA
 651 AGACAATGCC AATACCCAGG TTGCGGAAAA AGGCGTAGCC GATCTGCTCA
 701 AAGCGCGGCA CGGCGCGGAA GATTCTTCTA TGAACAACAG CGACAGCATC
 751 AGCGCAGTAG TCGAAGACAC CACCGGTACG ATGAAGCTGC GTATTTCTCT
 801 CATCGCGCTG ATTCATATGG TATGTCGCGG CATCGCGCTG ATGAACATCA
 851 TCGTGTGCTC CGTACCGAG CGACACAAG AATCGCGCAT ACGGATGCGA
 901 ATCGGCGCGC GCGCGCGCAA TATTTGCAG CAGTTTGTGA TTGACGGGT
 951 GTTAATCTGC GTCAATCGCG GTTTGGTCGG CGTGGGTGTT TCGCGCGCG
 1001 TCAGCCTCGT GTTCAATCAT TTGTAAACG ACTTCGCGAT GGACATTTC
 1051 GCCATCTCGC TCATCGGCGC GGTGCGCTGT TCGACCGGAA TCGGACATCG
 1101 GTTCGCTTT ATGCTGCCA ATAAAGCAG CAAACTCAAT CCGATAGACG
 1151 CATTGGCACA GGATTGA

This corresponds to the amino acid sequence <SEQ ID 534; ORF134-1>:

1 MSVQAVLAHK MRSLLTMLGI IIGIASVVSV VALNGNSQKK ILEDISSIGT
 51 NTISIFPGRG FGDRRSGRIR TLITIDAKII AKQSYVASAT PMSSSGGTLL
 101 YRNLDLTLASL YVGGEQYFDV RGLKLETGR LFDENVKEDA QVVVLDQNVK
 151 DLKFLADPL KTLTLERRP LTVIGMKKD ENAFNGSDVL MLMSPYTTVM
 201 HQITGESHTN SITVTKKDNA WVGVAEKLK DLKARSGTE DFMNNSDSI
 251 RQIVESTTGT MKLLISSIAL ISLVVVGIGV MNIMLVSVTE RTEIGIRMA
 301 IGARRGNILQ QFLTEAVLIC VIGGLVGVGL SAAVSLVENH FVTDFPMDIS
 351 AMSVIGAVAC STGIGIAFGF MPANKAKLN PIDALAQD*

Computer analysis of this amino acid sequence gave the following results:

Homology with the hypothetical protein o648 of *E. coli* (accession number AE000189)

ORF134 and o648 protein show 45% aa identity in 153aa overlap:

Orf134: 2 RHGTEDEFFMNSDXIRQIVESTTGTMKLXXXXXXXXXXXXXVGGIGVMNIMLVSVTERKEI 61
 55 RHG +DFF N D + + VE TT T+ + VVGIGVMNIMLVSVTERTEI
 o648: 496 RHGKDDFTWNMDGVLKTVKTRTLQLFLTLVAVISLVVGGIGVMNIMLVSVTERTEI 555
 Orf134: 62 GIRMAIGARRGNIXQFLIEAXXXXXXXXXXXXXXFNHFTDFPMDISAMSVI 121
 GIRMA+GAR ++ QQFLIEA F+ + + S + + + +
 o648: 556 GIRMAVGARASDVLLQFLIEAVLVCLVGGALGITLSLLIAFTLQLFLPGWEIGFSPALL 615

orf134: 122 GAVACSTGIGIAFGFMPANKAAKLNPIDALQD 154
 A CST GI FG++PA AA+L+P+DALA++
 o648: 616 LAFLCSTVTGILFGLFARNARLDPVDALARE 648

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF134 shows 98.7% identity over a 154aa overlap with an ORF (ORF134a) from strain A of *N.*

meningitidis:

```

10 orf134.pep      ARHGTEDFFMNSDXIRQIVESTGTMTKLL
orf134a      GESHTNSITVKKIKDNANTQVAEGLTDLKARHGTEDFFMNSDSIRQIVESTGTMTKLL
                210      220      230      240      250      260

15 orf134.pep      ISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMAIGARRGNIXQQFLIEAVLICVIGG
orf134a      ISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMAIGARRGNILQQFLIEAVLICVIGG
                270      280      290      300      310      320

20 orf134.pep      LVGVGLSAAVSLVFNHFVTDFPMDISAMSVIGAVACSTGIGIAFGFMPANKAAKLNPIDA
orf134a      LVGVGLSAAVSLVFNHFVTDFPMDISAMSVIGAVACSTGIGIAFGFMPANKAAKLNPIDA
                330      340      350      360      370      380

25 orf134.pep      LAQDX
orf134a      LAQDX
                IIII
30 orf134a      LAQDX

```

The complete length ORF134a nucleotide sequence <SEQ ID 535> is:

```

1  ATGTCGGTGC AAGCAGTATT GCGGCACAAA ATCGGTTGCG TTCTGACGAT
5  GCTCGGCATC ATCATCGGTA TCGCTTCGGT TGTCCTCGTC GTCGCATTGT
101 GCACACGGTTC GCAGAAAAAA ATCCCTGAAG ACATCAGTTC GATAGGAGCG
151 AACACCATCA GCATCTTCCC AGGGCGCGCG TTCGCGACA GCGCAGCGG
201 CAGGATTAAA ACCCTGACCA TAGACGACGC AAAAATCATC GCCAAACAAA
251 GCTACGTTGC TTCGCGACG CCGATGACTT CGAGCGCGCG CACGCTGACT
301 TACCGCAATA CCGACCTGAC CGCTTCTTTG TACGCTGTGG GCGAACAAATA
351 TTTGACAGTG CCGCGGCTGA AGCTGGAACG GGGCGCGCTG TTTGACGAAA
401 ACGATGTGAA AGAAGACGCG CAGGTCGTGC TCAATGACCA AAATGTCAAA
451 GACAAACTCT TTGCGGACTC GGATCCGTTG GGTAAACCCA TTTTGTTCAG
501 GAAACGCCCC TTGACCTGCA TCGGCGGTAT TGAAGAAGAC GAAACGCTT
551 TCGGCAATTC CGACGTGCTG ATGCTTTGTT CGCCCTATAC GAGCGTGATG
601 CACCAAAATCA CAGGCGAGAG CCACACCAAC TCCATCACC GCATCAATCA
45 651 AGCAAAATGCC AATACCCAGG TTGCGGAAAA AGGGCTGACC GATCTGCTCA
701 AAGCGCGGCA CCGCACGGAA GATTTCCTCA TGAACAACAG CGACAGCATC
751 AGGCAGATAG TCGAAGCAC CACCGGTACG ATGAAGCTGC TGATTTCCTC
801 CATCGCCCTG ATTTCAATTG TAGTCGGCGG CATCGGCTG ATGAACATCA
50 851 TGCTGTGTGC GGTACCGAG CGCACCAAAG AAATCGGCAT ACGGATGGCA
901 ATCGCGCGCG GCGCGGCGCA TATTTTGACG CAGTTTTTTG TTGAGCGCGG
951 GTTAATCTGC GTTCATCGCG GTTTGGTCGG GTGGGCTTTG TCGCGCGCGG
1001 TCAGCTCCGT GTTCAATCAT TTTGTAACCG ACTTCCGATG GGACATTTCC
1051 CGCATCTCGC TCATCGCGCG GTGCGCGCTG TCGACCGGAA TCGGCATCGC
55 1101 GTTCGCGCTT ATGCGCTGCC A7AAGCAGC CAAATCTCAAT CCGATAGATG
1151 CATTGCGCGA GGATTGA

```

This encodes a protein having amino acid sequence <SEQ ID 536>:

```

1  MSVQAVLAHK MRSLLTMLGI IIGIASVVSV VALGNGSQKK ILEDISSIGT
51 NTISIFFPGRG FGDRRSGRKI TLTIDDAKII AKQSYVASAT PMTSSGGTLT
60 71 YRNTDLTASL YGVGEQYFDV RGLKLETGR LFDENVKEDA QVVVIDQNVK
151 DKLFADSDPL GKTILFRKRP LTVIGVMKGD ENAFNGSDVL MLNFSPTTVM
201 HQITGESHTN SITVKIKDNA NTQVAEKGIL DLLKARHGTG DFFMNNSDSI
251 RQIVESTTGT MKLLISSIAL ISLVVGGIGV MNIMLVSVTE RTKEIGIRMA

```

301 IGARRGNILQ QFLIEAVLIC VIGGLVGVGL SAAVSLVFNH FVTDFFMDIS
351 AMSVIGAVAC STGIGIAFGF MPANKAAKLN PIDALAQD*

ORF134a and ORF134-1 show 100.0% identity in 388 aa overlap:

5	orf134a.pep	MSVQAVLAHKMRSLLTMLGIIIGIASVVSVALNGSGKKILEDISSIGTNTISIFPGRG	
	orf134-1	MSVQAVLAHKMRSLLTMLGIIIGIASVVSVALNGSGKKILEDISSIGTNTISIFPGRG	
	orf134a.pep	FGDRRSGRIKTLTIDDAKIIAKQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEQYFDV	
10	orf134-1	FGDRRSGRIKTLTIDDAKIIAKQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEQYFDV	
	orf134a.pep	RGLKLETGRLFDENDVKEDAQVVVIDQNVKDKLFADSDPLGKTLFRRKRLTVIGVMKKD	
	orf134-1	RGLKLETGRLFDENDVKEDAQVVVIDQNVKDKLFADSDPLGKTLFRRKRLTVIGVMKKD	
15	orf134a.pep	ENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNANTQVAEKGLTDLKARHGTE	
	orf134-1	ENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNANTQVAEKGLTDLKARHGTE	
20	orf134a.pep	DFPMNNSDSIRQIVESTTGTMKLLISSIALISLVGGIGVMNIMLVSVTERTKEIGIRMA	
	orf134-1	DFPMNNSDSIRQIVESTTGTMKLLISSIALISLVGGIGVMNIMLVSVTERTKEIGIRMA	
	orf134a.pep	IGARRGNILQQFLIEAVLICVIGGLVGVGLSAAVSLVFNHFVTDFFMDISAMSVIGAVAC	
25	orf134-1	IGARRGNILQQFLIEAVLICVIGGLVGVGLSAAVSLVFNHFVTDFFMDISAMSVIGAVAC	
	orf134a.pep	STGIGIAFGFMPANKAAKLNPIDALAQDX	
30	orf134-1	STGIGIAFGFMPANKAAKLNPIDALAQDX	

Homology with a predicted ORF from *N.gonorrhoeae*

ORF134 shows 96.8% identity over a 154aa overlap with a predicted ORF (ORF134.ng) from *N. gonorrhoeae*:

35	orf134.pep	ARHGTEDDFMNNSDXIRQIVESTTGTMKLL	30
	orf134ng	GESHTNSITVKIKDNANTRVAEKGIAELLKARHGTEDDFMNNSDSIRQIVESTTGTMKLL	264
40	orf134.pep	ISSIALISLVGGIGVMNIMLVSVTERTKEIGIRMAIGARRGNIXQQFLIEAVLICVIGG	90
	orf134ng	ISSIALISLVGGIGVMNIMLVSVTERTKEIGIRMAIGARRGNILQQFLIEAVLICIIIG	324
	orf134.pep	LVGVGLSAAVSLVFNHFVTDFFMDISAMSVIGAVACSTGIGIAFGFMPANKAAKLNPIDA	150
45	orf134ng	LVGVGLSAAVSLVFNHFVTDFFMDISAAVIGAVACSTGIGIAFGFMPANKAAKLNPIDA	384
	orf134.pep	LAQD 154	
	orf134ng	LAQD 388	

50 The complete length ORF134ng nucleotide sequence <SEQ ID 537> is:

1	ATGTCGGTGC	AAGCAGTATT	GCGGCACAAA	ATGCGTTGCG	TTCTGACCAT
51	GCTCGGCATC	ATCATCGGTA	TGCTTCGGT	GTCTCGGTC	GTCGCGCTGG
101	GCAACGGTTC	GCAGAAAJAA	ATCCTCGAAG	ACATCAGTTC	GATCGGGACG
151	AACACCATCA	GCATCTTCCC	CGGGCGCGCG	TTGGCGACAC	GCGCAGCGCG
55	201	CAAAATCAAA	ACCCGTGACCA	TAGACGACGC	AAAAATCATC
	251	GCTACGTGTC	CTCCGCCACG	CCCATGACTT	CGAGCGGCGG
	301	TACCCGCAATA	CCGACCTGAC	GCCTTCTTTG	TACGGTGTGG
	351	TTTCGACGTG	CGCGGGCTGA	AGCTGGAAAC	GGGCGGCGTG
	401	ACGATGTGAA	AGAGACGCGC	CAAGTCGTGC	TCATCGACCA
60	451	GACAAACTCT	TTGCGGACCT	GGATCCGTTG	GGTAAACCA
	501	GAAACGCCCC	TTGACGCTCA	TCGCGGTGAT	GAAAAAGAC
	551	TCGCGCAATC	CGACGTGCTG	ATGCTTTGGT	GCGCCATATC

-332-

5 601 CACCAAAATCA CAGGCGAGAG CCACACCAAC TCGATCACCG TCAGAAATCAA
 651 AGACAATGCC AATACCGGGG TTGCGGAAAA AGGGCTGGCC GAGCTGCTCA
 701 AAGCAGCGCA CGGCGCGGAA GACTCTTTTA TGAACAACAG CGACGAGATC
 751 AGGCAGATGG TCGAAGCAC CACCGGTACG ATGAAGCTGC TGATTTCTCT
 801 CATCGCCCTG ATTTCATTGG TAGTCGGCGG CATCGGTGTG ATGAACATTA
 851 TCGTGTGTGC CTATACGAG CGACACAG AATCGGAT ACGSATGCA
 901 ATCGGCGCGC GCGCGGCA TATTTTGAC GAGTTTGA TTGAGCGGCT
 951 GTTAATCTGC ATCATCGGAG GCTTGTTCGG CCGTAGTTTC TCGCGCCCG
 10 1001 TCAGCCTCGT GTTCAATCAT TTTGTACCG ATTTCCGAT GGACATTTCG
 1051 GCGGCATCCG TTATCGGGGC GGTGCGCTGT TCGACCGGAA TCGGCATCGC
 1101 GTTCGGCTTT ATGCTTGCCA ATAAGGCAG CAACTCAAT CCGATAGATG
 1151 CATTGGCGCA GGATTGA

This encodes a protein having amino acid sequence <SEQ ID 538>:

15 1 MSVQAVLAHK MRSLLTMLGI IIGIASVVSVALNGSQKK ILEDISSMGT
 51 NTISIFFGRG FGDRRSRGKIK TLTIDDAKII AKQSYVASAT PMTSSGGTLT
 101 YRNTDLTASL YGVGEQYFDV RGLKLETGRIL FDENDVKEDA QVVVIDQNVK
 151 DKLFADSDPL GKTLIFKRKP LTVIGVMKKD ENAFGNSDVL MLWSPYTTVM
 201 HQITGESHTN SITVKIKDNA NTRVAEKGLE LLKARHGT E DFFMNSDSI
 251 RQMVESPTGT MKLLISSIAL ISLVGGIGV MNIMLVSVTE RTKEIGIRMA
 301 IGARRGNILQ QELIEAVLIC IIGSLVGVGL SAARVSVFNH FVTDFPMDIS
 351 AASVIGAVAC STGIGIAFGF MFANKAKLIN PIDALAQD*

ORF134ng and ORF134-1 show 97.9% identity in 388 aa overlap:

25 orf134ng MSVQAVLAHKMRSLLTMLGIIIGIASVVSVALNGSQKKILEDISMGTNTISIFPGRG
 orf134-1 MSVQAVLAHKMRSLLTMLGIIIGIASVVSVALNGSQKKILEDISIGTNTISIFPGRG
 30 orf134ng FGDRRSRGKIKTLTIDDAKIIAKQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEQYFDV
 orf134-1 FGDRRSRGKIKTLTIDDAKIIAKQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEQYFDV
 35 orf134ng RGLKLETGRILFDENDVKEDAQVVVIDQNVKDKLFADSDPLGKTLIFKRKPLTVIGVMKKD
 orf134-1 RGLKLETGRILFDENDVKEDAQVVVIDQNVKDKLFADSDPLGKTLIFKRKPLTVIGVMKKD
 40 orf134ng ENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNA NTRVAEKGLELLKARHGT
 orf134-1 ENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNA NTRVAEKGLELLKARHGT
 45 orf134ng DFFMNSDSIRQMVESPTGTMTKLLISSIALISLVGGIGVMNIMLVSVTERTEIGIRMA
 orf134-1 DFFMNSDSIRQIVESTTGTMTKLLISSIALISLVGGIGVMNIMLVSVTERTEIGIRMA
 orf134ng IGARRGNILQQFLIEAVLICIIIGSLVGVGLSAARVSVFNHFVTDFPMDIS AASVIGAVAC
 50 orf134-1 IGARRGNILQQFLIEAVLICIIIGSLVGVGLSAARVSVFNHFVTDFPMDIS AASVIGAVAC
 orf134ng STGIGIAFGFMFANKAKLINPIDALAQDX
 orf134-1 STGIGIAFGFMFANKAKLINPIDALAQDX

ORF134ng also shows homology to an *E.coli* ABC transporter:

sp|P75831|YBJZ_ECOLI HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN YBJZ >gis
 (AE000189) o648; similar to YBBA_HAEN SW: P45247 [Escherichia coli] length =
 648
 55 Score = 297 bits (753), Expect = 6e-80
 Identities = 162/389 (41%), Positives = 230/389 (58%), Gaps = 1/389 (0%)
 Query: 1 MSVQAVLAHKMRSLLTMLIXXXXXXXXXXXLNGSQKKILEDISMGTNTISIFPGRG 60
 M+ *A+ A+KMR+LLTML *G+ *++ +L DI S+GNTII ++PG+
 60 Sbjct: 260 MAWRALAANKMRLLTMLGIIIGIASVVSIVVVGDAKQMLADIRISGTNTIDVYGGK 319
 Query: 61 FGDRRSRGKIKTLTIDDAKIIAKQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEQYFDV 120
 FGD +L DD I KQ +VASATP S L Y N D+ AS GV YF+V
 Sbjct: 320 FGDDDPQYQALKYDDLIAIQKQPVASATPAVSGNRLRLRYNNVDVAASANGVSGDYFNV 379

Query: 121 RGLKLETGRLFDENDVKEDAQVVVIDQNVKDKLFAD-SDPLGKTLILFRKRPLTVIGVMKK 179
 G+ G F+ + + AQVVVD N + +LF P VIG + +
 Sbjct: 380 YGMTFSEGTENFQEQNLNGRAQVVVLDSNTRQLFPHKADVGEVLVGNMFPARVIGVAEE 439

Query: 180 DENAPGNSDVLMLNSPYTTVMHQITGESHTNSITVKIKDNANTRVAEKGAEALLKARHGT 239
 ++ PG+S VL +W PY+T+ ++ G+S NSITV++K+ ++ AE+ L LL RHG
 Sbjct: 440 KQSMFGSSKVLRLVWLYSTMSGRVMGQSWLNSITVRVEKGFDSAEEQQLTRLLSLRHKG 499

Query: 240 EDFFMNNSDSIRQMVSTTGTMKXXXXXXXXXXXXVVGIGVMNIMLVSVTERTKEIGIRM 299
 +DEF N D + + VE TT T++ VVGIGVMNIMLVSVTERT+EIGIRM
 Sbjct: 500 KDFFTNMNDGVLKTVEKTTTRLQLFLFLVAIVSLVVGIGVMNIMLVSVTERTREIGIRM 559

Query: 300 AIGARRGNILQQFLIEXXXXXXXXXXXXXXXXXFNHFVTFDFPMDISAAVIGAVA 359
 A+GAR ++LQQFLIE F+ + + S +++ A
 Sbjct: 560 AVGARASDVLQQFLIEAVLVCLVGGALGITLSLLIAFTLQLFLPGMEIGFSPFALLAFL 619

Query: 360 CSTGIGIAGFGMPANKAAKLNPIDALAQD 388
 CST GI FG++PA AA+L+P+DALA++
 Sbjct: 620 CSTVTGILFGWLPAARNAARLDFVDALARE 648

Based on this analysis, including the presence of the leader peptide and transmembrane regions in the gonococcal protein, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 65

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 539>:

```

1 ..GGGACGGGAG CGATGCTGCT GCTGTTTAC GCGSTAACGA T.CTGCCTTT
51 GGCCACTGGC GTTACCTCGA GTTACACCTC CTGATTATTT TTGGCGGTAT
101 TTCTCTTCCT GATTTTGAAG GAAACGATTT CCGTTTACAC GCAGGCGGTG
151 CTGCTCCTTG GTTTTGC CGG CGTGATATTG CTGCTTAATC CCTCGTCCG
201 CAGCGGTGAG GAAACGGGG CACTCGCGGG GCTCGGGGG GCGCGATGT
251 CCGCTGGGCG GTATTGAGAA GTGCGGGAAC TGCTCTTGGC GGGCGAACCC
301 GGCTGGGCGG TCGTGTTTTA CCTTCCGCTG ACAGGTGTGG CGATGTGCTG
351 GCTTTGGGCG ACGCTGACCG GCTGCGAAC CCGTCCCTT CCATCGGCG
401 TTATATCTGC GTGCTCGGCG GTGTCGCGCG TCATTGCGCA ACTGTCGATG
451 ACGCGGCGCT ACAAAAGTCG CGACAAATTC ACGGTTGCCT GCCTTTCTTA
501 TATGACGCTC GTTTTTCCTG CTCTGTCTGC CGCATTTTTC CTGGGCGAAG
551 AGCTTTTCTG GCAGGAAATA CTCGATATGT GCATCATCAT CCGTACCGGT
601 ATTTTGA
  
```

This corresponds to the amino acid sequence <SEQ ID 540; ORF135>:

```

1 ..GTGAMLLFF AVTILPLATG VTLSTYSIF LAVFSFLILK ERISVYTOAV
51 LLLGFAGVVL LLNFSFRSGO ETAAALAGLAG GAMSQWAYLK VRELSLAGEP
101 GWRVVFYLSV TGVAMSSVWA TLTGWHILSF PSAVYLSIG VSAIAQLISM
151 TRAYKVGDKF TVASLSYMTV VFSALSAFF LGELFLWQEI LGMCIIISAV
201 F*
  
```

Further work revealed the complete nucleotide sequence <SEQ ID 541>:

```

1 ATGGATACCG CAAAAAAGA CATTTAGGA TCGGCTCGA TGCTGGTGGC
51 GCGGCGCTCG TTTACCATTA TGAACGTATT GATTAAAGG GCATCGGCA
101 AATTTCGCCCT CGGACGGCGC GAATTGCTCT TTTGGCGCAT GGTGTTTTCA
151 ACCGTTCGGC TCGGGGCTGC CGCGGATATT GCTCGGCAAC MCTTCGGCAC
201 GCCCATATGG AAAAACCACT TAAACCGCAC TATGGTCCGG ACGGGGGCGA
251 TGCTGCTGCT GTTTTACGCG GTAACGCATC TGCCCTTGGC CACTGGCGGT
301 ACCCTGAGTT ACACCTCGTC GATTTTITTT GCGGTATTTT CTTCTCTGAT
351 TTTGAAAGAA CGGATTTCCG TTTACACGCA GCGCGTGTCT CTCTTGGT
401 TTGCGCGCGT GGTATTGCTG CTTAATCCCT CGTTCGCGAG CGGTCAAGAA
451 ACGGCGGCGC TCGCGGGGCT GCGGGGCGCG GCGATGTCCG GCTGGGCGTA
501 TTTGAAAGTG CCGCAACTGT CTTTGGCGGG CGAACCCGCG TGCGCGCTCG
551 TGTTTTACCT TTCGTGACA GTGTGGCGCA TGTCTGCTGT TTGGGCGAGC
  
```

5
 601 CTGACCGGCT GGCACACCCCT GTCCTTTCCA TCGGCAGTTT ATCTGTCGTG
 651 CATCGGCGTG TCCGCGCTGA TTGCCCAACT GTCGATGACG CGCGCTACAA
 701 AAGTCGGCGA CAAATTCACG GTTGCTCTGC TTCTCTATAT GACGCTCGTT
 751 TTTTCGCTC TGCTCGCGC ATTTCTCTGC GCGCAAGAGC TTTCTGSCA
 801 GCAAACTACT GGTATGTGCA TCATCATCTC CAGCGGTATT TTGAGACACA
 851 TCCGCCCCAC TGCTTCAAAA CAGCGCTGCG AATCCCTGTT CCGCCAAAGA
 901 TAA

This corresponds to the amino acid sequence <SEQ ID 542; ORF135-1>:

10
 1 MDTAKKDI LG SGWMLVAAC FTIMNVLIKE ASAKFALGSG ELVFWRMFLS
 51 TVALGAAAVL RRDxFRTPHW KNHLNRSVMG TGAMLLLFYA VTHLPLATGV
 101 TLSYTSISIFL AVFSFLILKE RISVYTOAVL ILGFAGVVLN LNPFSRSGOE
 151 TAALAGLAGG AMSGWAYLKV RELSLAGEPG WRVVFYLSVT GVAMSSVWAT
 201 LTGWHTLSFP SAVYLSICGV SALIAQLSMT RAYKVGDKFT VASLSYMTVV
 15
 251 FSALSAFFL GEELFWQEIL GMCIIILSGI LSSIRPTAFK QRLQSLFRQR
 301 *

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF135 shows 99.0% identity over a 197aa overlap with an ORF (ORF135a) from strain A of *N. meningitidis*:

20
 orf135.pep 10 20 30
 GTGAMLLLFYAVTILPLATGVTLSSYTSIF
 orf135a 50 60 70 80 90 100
 STVALGAAAVLRRDFTERTPHWKNHLNRSVMGTGAMLLLFYAVTILPLATGVTLSSYTSIF
 25
 orf135.pep 40 50 60 70 80 90
 LAVFSFLILKERISVYTOAVLLLGFGAGVVLNLPFSRSGQETAALAGLAGGAMSGWAYLK
 orf135a 110 120 130 140 150 160
 LAVFSFLILKERISVYTOAVLLLGFGAGVVLNLPFSRSGQETAALAGLAGGAMSGWAYLK
 30
 orf135.pep 100 110 120 130 140 150
 VRELSLAGEPGWRVVFYLSVTGVAMSSVWATLTGWHTLSFPASVYLSICGVSAIAQLSM
 orf135a 170 180 190 200 210 220
 VRELSLAGEPGWRVVFYLSVTGVAMSSVWATLTGWHTLSFPASVYLSICGVSAIAQLSM
 35
 orf135.pep 160 170 180 190 200
 TRAYKVGDKFTVASLSYMTVVSALSAFFLGEELFWQEILGMCIIISAVFK
 40
 orf135a 230 240 250 260 270 280
 TRAYKVGDKFTVASLSYMTVVSALSAFFLAEELFWQEILGMCIIILSGILSSIRPTAF
 45
 orf135a 290 300
 KQRLQSLFRQRX

The complete length ORF135a nucleotide sequence <SEQ ID 543> is:

50
 1 ATGGATACCG CAAAAAAGA CATTTTAGGA TCGGCGTGGA TGCTGGTGGC
 51 GCGCGGCGTCG TTTACCATTA TGAACGTATT GATTAAAGAG GCATCGGCAG
 101 AATTTCGCCCT CGGCAGCGCG GAATTGTGCT TTTGCGCGCAT GCTGTTTTCAC
 151 ACCGTTGCGCG TCGGGGCTCG CGCCGTATTG CGTCGGGACA CCTTCGCGAC
 201 GCCCCATTGG AAAAAACCACT TAAACCGCAG TATGTCGCGG ACGGGGGCGCA
 251 TGCTGCTGCT GTTTTACGCG GTAACGCATC TGCCCTTTGGC CACCGCGCTT
 301 ACCCTGAGTT ACACCTCTGC GATTTTTTCG GCGGTATTTT CTTCTCGTAT
 351 TTTCAAGAGA CGATTCTCG TTACACGCA GCGCGTGCTG CTCTTGGTGT
 401 TTGCGCGCGCT GGTATTCTCG CTTTACGCG CTTTCCGCGC CGGTGAGGAA
 451 ACGGCGGCGAC TCGCGGCGCT GCGGGGCGCG CGGATGTCGG CTGCGGCGTA
 501 TTTGAAAGTG CGCGAACTGT CTTTGGCGGG CGAACCCGCG TGGCGCGTCG
 551 TGTTTTACCT TTCCTGACAA GGTGTGGCGA TGTCATCGGT TTGGCGGACG
 601 CTGACCGGCT GGCACACCCCT GTCTTTTCCA TCGGCAGTTT ATCTGTCGTG

5 651 CATCGGCGTG TCCGCGCTGA TTGCCCAACT GTCGATGACG CGCGCCTACA
701 AAGTCGGCGA CAATTCACG GTTGCTCGC TTTCCTATAT GACCGTCGTT
751 TTTTCCGCTC TGCTCGCGCG ATTTTCTCTG GCCGAGAGAGC TTTTCTGGCA
801 GGAATACTC GGTATGTGCA TCATCATCTC CAGCGGTATT TTGAGCAGCA
851 TCCGCCCCAC TGCCTTCAA CAGCGGCTGC AATCCCTGTT CGGCCAAGA
901 TAA

This encodes a protein having amino acid sequence <SEQ ID 544>:

10 1 MDTAKKDILG SGWMLVAAAC FTIMNVLIKE ASAKFALGSG ELVFWRMFLVS
2 TVALGAAAVL RRDTERTPHW KHNHNRSMVG TGAMLLLFYA VTHLPLATGV
101 TLSTSSIFL AVFSFLIKE RISVYTOAVL LIGFAGVVL LNFSFRSGQE
151 TRALAGLAGG AMSGWAYLKV RELSLAGEFG WRVVFYLSVT GVAMSSVWAT
201 LTGWHTLSFP SAVYLSGIGV SALIAQLSMT RAYKVGDKFT VASLSYMTVW
251 FSALSAFFEL AEELFWQEL GMCTIILSGI LSSIRPTAFK QRLQSLFRQR
301 *

15 ORF135a and ORF135-1 show 99.3% identity in 300 aa overlap:

orf135a.pep MDTAKKDILGSGWMLVAAACFTIMNVLIKEASAKFALGSGELVFWRMFLSTVALGAAAVL
orf135-1 MDTAKKDILGSGWMLVAAACFTIMNVLIKEASAKFALGSGELVFWRMFLSTVALGAAAVL
20 orf135a.pep RRDTERTPHWKHNHNRSMVGTGAMLLLFYAVTHLPLATGVTLSTSSIFLAVFSFLIKE
orf135-1 RRDTERTPHWKHNHNRSMVGTGAMLLLFYAVTHLPLATGVTLSTSSIFLAVFSFLIKE
25 orf135a.pep RISVYTOAVLLLGAGVGVLLNPSFRSGQETAALAGLAGGAMSGWAYLKVRELSLAGEFG
orf135-1 RISVYTOAVLLLGAGVGVLLNPSFRSGQETAALAGLAGGAMSGWAYLKVRELSLAGEFG
30 orf135a.pep WRVVFYLSVTGVAMSSVWATLTGWHTLSFSAVYLSGIGVSALIAQLSMTRAYKVGDKFT
orf135-1 WRVVFYLSVTGVAMSSVWATLTGWHTLSFSAVYLSGIGVSALIAQLSMTRAYKVGDKFT
35 orf135a.pep VASLSYMTVVSALSAFFLAELFWQELGMCTIILSGILSSIRPTAFKQRLQSLFRQR
orf135-1 VASLSYMTVVSALSAFFLGEELFWQELGMCTIILSGILSSIRPTAFKQRLQSLFRQR

Homology with a predicted ORF from *N.gonorrhoeae*

ORF135 shows 97% identity over a 201aa overlap with a predicted ORF (ORF135ng) from *N.gonorrhoeae*:

40 orf135.pep GTGAMLLLFYAVTXLPLATGVTLSTSSIF 30
orf135ng STVTLGAAAVLRDTERTPHWKHNHNRSMVGTGAMLLLFYAVTHLPLTTGVTLSYSSIF 335
orf135.pep LAVFSFLIKERISVYTOAVLLLGAGVGVLLNPSFRSGQETAALAGLAGGAMSGWAYLK 90
45 orf135ng LAVFSFLIKERISVYTOAVLLLGAGVGVLLNPSFRSGQETAALAGLAGGAMSGWAYLK 395
orf135.pep VRELSLAGEPGWRVVFYLSVTGVAMSSVWATLTGWHTLSFSAVYLSGIGVSALIAQLSM 150
50 orf135ng VRELSLAGEPGWRVVFYLSATGVAMSSVWATLTGWHTLSFSAVYLSGIGVSALIAQLSM 455
orf135.pep TRAYKVGDKFTVASLSYMTVVSALSAFFLGEELFWQELGMCTIISAVF 201
orf135ng TRAYKVGDKFTVASLSYMTVVSALSAFFLGEELFWQELGMCTIISAAF 506

An ORF135ng nucleotide sequence <SEQ ID 545> was predicted to encode a protein having amino acid sequence <SEQ ID 546>:

1 MPSEKAFRRH LRTASFQGLH LHHFHQKVGK CGIIGFGIHI FPTLLPAQGW
51 ILDIQLGLFR IDFAALAVYR RTQVDFIHTV IDGIASDAQF SEVVQILRL
101 NLGHFTDTHL IAQARRFIAD FGNIRPMRGR EAKTFCRCFR FDGIDGIND

151 FROCGHINRL APGKDCRNKG RDKVFFHTRH YNQVCLEKTN CSARKIKFRH
 201 QKQARTHSTS LAARFTIRS LSORPFMDTA KKDILGSGWM LVAAACFTVM
 251 NVLKEASAK FALGSGELVF WRMLFSTVTL GAAAVLRDT FRTPHWKHL
 301 NRSVMGTGAM LLLFYAVTHL PLTTGVTLSY TSSIFLAVFS FLILKERISV
 351 YTOQVLLLG FAGVLLLNPS FRSGQEPAL AGLAGGAMSG WAYLKVRELS
 401 LAGEPGNRVV FYLSATGVAM SSVWATLTGW HTLSFESAVY LSGIGVSALI
 451 AQLSMTRAYK VGKFTVASL SYMTVVFSAI SAAFLGEEL FWQELILMCII
 501 IIIAAF*

Further work revealed the following gonococcal sequence <SEQ ID 547>:

1 ATGGATACCG CAAAAAAGA CATTTTAGGA TCGGGCTGGA TGCTGGTGGC
 51 GCGGCGCTGC TTCACCGTTA TGAACGTATT GATTAAAGAG GCATCGGCAA
 101 AATTTCGCCCT CGGCAGCGCG GAATTGGTCT TTTGGCGCAT GCTGTTTCA
 151 ACCGTTACGC TCGGTGCTGC CGCGGTATTG CGGCGCGACA CTTTCGCGAC
 201 GCCCATTTGG AAAAACCAGT TAAACGCGAG TATGTTCGGG ACGGGGCGCA
 251 TGCCTGCTGCT GTTTTACGCG GTAACGCATC TGCCCTTGAC AACCGCGGTT
 301 ACCTGAGATT ACACCTTCGC GATTITTTtG GCGGTATTTT CTTCTCGAT
 351 TTTGAAAGAA CGGATTTCGC TTTACACGCA GCGGSGCTG CTTCTGGTT
 401 TTGCGCGCGT GGTATTGCTG CTTAATCCCT CGTTCGCGAG CGGTACGAA
 451 CCGGCGCGAC TCGCGCGCGT GCGGCGCGG CGCATGTCGG GCTGGCGTA
 501 TTGGAAGTG CGGCGACTGT CTTTGGCGGG CGAACCCGGC TGGCGCGTCG
 551 TGTTTACCT TTCCGCAACC GCGGTGGCGA TGTCGTGggt ttgggagag
 601 Ctgacccgct ggCACaccct GTCCTTcca tggcgagttt ATCgtCGGG
 651 CATCGGCGGT tccgcgCtgA TTGCCCAact GtcgatAgc cGCGctaca
 701 aaGTGCGGGA CAAATTCACG GTTGCGCTCG tttcctaTat gaccgctGTC
 751 TTTTCCGCCC TGCTCGCGCG ATTTTITCTG ggcgaagagc tttTctggCA
 801 GGAATACTC GGTATGTGCA TCATTatccT CAGCGCATT TTGAGCAGCA
 851 TCGCGCCCAT TGCCTTCAAA CAGCGGCTGC AAGCCCTCTT CCGCCAAAGA
 901 TAA

This corresponds to the amino acid sequence <SEQ ID 548; ORF135ng-1>:

1 MDTAKKDILG SGWMLVAAAC FTMNVNLIKE ASAKFALGSG ELVFWRMFLS
 51 TVTLGAAAVL RRDTRTPHW KNLNLRSMVG TGAMLLLFYA VTHLPLITGV
 101 TLSYSSIFL AVFSFLILKE RISVYTOAVL LLGFAGVVLL LNPSFRSGQE
 151 PAALAGLAGG AMSGWAYIKV RELSLAGEPG WRVVFVLSAT GVAMSSVWAT
 201 LTGWHTLSFP SAVYLSGIGV SALLAQLSMT RAYKVSDEKT VASLSYMTV
 251 FSALSAFFL GEELFWQELI GWCIIILSGI LSSIRPIAFK QRLQALFRQR
 301 *

ORF135ng-1 and ORF135-1 show 97.0% identity in 300 aa overlap:

orf135ng-1.pep MDTAKKDILGSGWMLVAAACFTMVNVLKEASAKFALGSGELVFWRMLFSTVTLGAAAVL
 40 orf135-1 MDTAKKDILGSGWMLVAAACFTMVNVLKEASAKFALGSGELVFWRMLFSTVTLGAAAVL
 orf135ng-1.pep RRDTRTPHWKNLNRSMVGTGAMLLLFYAVTHLPLTTGVTLSYSSIFLAVFSFLILKE
 45 orf135-1 RRDTRTPHWKNLNRSMVGTGAMLLLFYAVTHLPLATGVTLSYSSIFLAVFSFLILKE
 orf135ng-1.pep RISVYTOAVLLLGFAGVVLLLNPSFRSGQEPALAGLAGGAMSGWAYIKVRELSAGEPG
 orf135-1 RISVYTOAVLLLGFAGVVLLLNPSFRSGQEPALAGLAGGAMSGWAYIKVRELSAGEPG
 50 orf135ng-1.pep WRVVFVLSATGVAMSSVWATLTGWHTLSFSAVYLSGIGVSALLAQLSMTRAYKVGDKFT
 orf135-1 WRVVFVLSATGVAMSSVWATLTGWHTLSFSAVYLSGIGVSALLAQLSMTRAYKVGDKFT
 orf135ng-1.pep VASLSYMTVVFSAALSAFFLGEELFWQELILGMCIILSGILSSIRPIAFKORLQALFRQR
 55 orf135-1 VASLSYMTVVFSAALSAFFLGEELFWQELILGMCIILSGILSSIRPIAFKORLQALFRQR

Based on this analysis, including the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.


```

      60      70      80      90      100      110      119
orf136.pep PCGIVFGALLERHLP AHCLYGKAAVGDVAHEHPVADVNNRANAFALFDIGQFAGFIVQ
5 orf136a  |||||:||||| :|||||:|||||:|||||:|||||:|||||:|||||
      70      80      90      100      110      120
orf136.pep PCGIVFGTLLERHFXSTHCLYGKAAVGNVAHEHPVADVNNRANAFALFDIGQFAGFIVQ
10 orf136a  |||||:|||||:|||||:|||||:|||||:|||||:|||||
      120      130      140      150      160      170      179
orf136.pep HTVNIKTVKINIVDPHMFANFAVFAVLEKRDFDHGKIQGGNNAAFPKKLAPKIFECFTG
15 orf136a  |||||:|||||:|||||:|||||:|||||:|||||:|||||
      130      140      150      160      170      180
orf136a  HAINVKTVKINIVDPHMFANFAVFAVLEKRALTMAKSKXXMMRRRSQKSKRYKLVNLA

      180      190      200      210      220      230
orf136.pep AFVGTVYRFVCLFYIINDGIAHH---SAPQVRVYLFAPYCGFLPSASDSLKSSXXSEX
15 orf136a  :||:| :|:| :|||:|||||:|||||:|||||:|||||
      190      200      210      220      230
orf136a  R---SPARFTGLSACSTXXMTESPIISAPQVRVYLFAPYCGFLPSASDSLKSSKYSEX

```

The complete length ORF136a nucleotide sequence <SEQ ID 553> is:

```

20 1 ATGATGAAGC GGCCTATAGC CGTCTTCGTC CTGCTCATGC AGAAAATCCG
51 GATTTCGCGA CAACGTGTGC CGAAAATCGT CAATACAGTT CCGGCACATC
101 CGATGCTCTT CCAGATNTTC GSGATGTTCT TTTCCTCATC ACACACAGCAA
151 TACCTGCCCG GGATCGCCGA AATCGATTCC CCAATCGCGCA TCGTGTTCGG
201 TACGCTCCTC TTCGGTCATC NGTCCACGCA TGCCCTGTAT GGTAAAGCCG
25 251 CCGTAGGGAA TGCCGTTCGA CACGAACATC CAGTCGTGTA TGTGCTCAAC
301 CGGAACGCAG ACGCTTTCGC CTTGTTCGAC ATTGGTCAGT TCGCCGGGTT
351 CATTTGTTCA GACGCCATAA ATGTAAGAAC CGTCAAATAA AATATCGCTG
401 ATCCACATAT GTTCGCAAAAT TTCGCCNTCT TCGCCGCTCT GGAAAAGAGG
30 451 GCTTTGACCA TGGCAAAATC TAAGGNGNNA NGATCGGGCG GCGCTTCCCA
501 AAAAAGCTCG GCGCAAAATC ATTTGAATGT TTTCCGGGCG GCTTCGCGCG
551 CACGGTTTAC CGGTTTGCTC GCCTGTCTCA CATATAAAT GACGGAATCG
601 CCCATCATAT CTGCTCCTCA ACGTGTACGG TATCTGTTTG CACCTTACTG
651 CGGCTTCTCG CTTTCGGCAT CCGATTCCGA TTTGAAAGT TCCAATATAT
701 CGGAATAG

```

This encodes a protein having amino acid sequence <SEQ ID 554>:

```

35 1 MMKRRIAVFV LLMQKIRILG QLLPKIVNTV PAHRMLFQXF GMFFFIHQV
51 YLPGIAEIDS PCGIVFGTLL FRHXSTHCLY GAAVGNVA HEHPVADVNN
101 RNANAFALFD IGQFAGFIVQ HAINVKTVKI NIVDPHMFAN FAXFAVLEKR
151 ALTMAKSKXX MMRRRSQKSS RKQYLVNLA RSPARTGLS ACST**MTES
40 201 PIISAPQVRV YLFAPYCGFL PSASDSLKS SKYSE*

```

ORF136a and ORF136-1 show 73.1% identity in 238 aa overlap:

```

      10      20      30      40      50      60
orf136a.pep MMKRRIAVFVLLMQKIRILGQLLPKIVNTVPAHRMLFQXFGMFFFIHQQYLPGLIAEIDS
45 orf136-1  |||||:|||||:|||||:|||||:|||||:|||||:|||||
      10      20      30      40      50      60
orf136a.pep MMKRRIAVFVLPQIIRVLGQLLPKIVNTVPAHRMLFQIFGMFFFIHQQYLPGLIAEIDS
      10      20      30      40      50      60
orf136-1  |||||:|||||:|||||:|||||:|||||:|||||:|||||
      70      80      90      100      110      120
orf136a.pep PCGIVFGTLLERHFXSTHCLYGKAAVGNVAHEHPVADVNNRANAFALFDIGQFAGFIVQ
50 orf136-1  |||||:|||||:|||||:|||||:|||||:|||||:|||||
      70      80      90      100      110      120
orf136-1  PCGIVFGALLERHLP AHCLYGKAAVGDVAHEHPVADVNNRANAFALFDIGQFAGFIVQ
      130      140      150      160      170      180
orf136a.pep HAINVKTVKINIVDPHMFANFAVFAVLEKRDFDHGKIQGGNNAAFPKKLAPKIFECFTG
55 orf136-1  |||||:|||||:|||||:|||||:|||||:|||||:|||||
      130      140      150      160      170      180
orf136-1  HTVNIKTVKINIVDPHMFANFAVFAVLEKRDFDHGKIQGGNNAAFPKKLAPKIFECFTG
      190      200      210      220      230
orf136a.pep R---SPARFTGLSACSTXXMTESPIISAPQVRVYLFAPYCGFLPSASDSLKSSKYSEX
60 orf136-1  :||:| :|:| :|||:|||||:|||||:|||||:|||||
      190      200      210      220      230
orf136-1  AFVGTVYRFVCLFYIINDGIAHH---SAPQVRVYLFAPYCGFLPSASDSLKSSKYSEX

```

190 200 210 220 230

Homology with a predicted ORF from *N.gonorrhoeae*

ORF136 shows 92.3% identity over a 234aa overlap with a predicted ORF (ORF136ng) from

5 *N.gonorrhoeae*:

```

orfl36.pep      MKRRIAVFVLFPQIIIRVLGQLLPKIVNTVPAHRMLFQIFGMFFFFFIHQQYLPGIAEIDS 59
orfl36ng      MMKRIIAVFVLLMQKIRILGQLLPKIVNTVPAHRMLFQIFGMFFFFFIHRLQYLPGIAEIDS 60
10 orfl36.pep      PCGIVFGALLFRHLPAHCLYGKAAVGDAVAHEHPVADVNRNANAFALFDIGQSAGFIVQ 119
orfl36ng      PGGIVFGTLLFRHLSAHCLYGKAAVGDAVAHEHPVADVNRNANAFALFDIGQSAGFIVQ 120
15 orfl36.pep      HTVNIKTVKINIVDPHMFANFAVFAVLEKRDFDHGKIQQGNNAAPFKKLAPKIFECFTG 179
orfl36ng      HTVNIKTVKINIVDPHMFANFAVFAVLEKRDFDHGKIQQGNNAAPFKKLAPKIFECFTG 180
orfl36.pep      AFVGTVYRVFVCLFYIINDGIAHHSAPQRVRYLFAPYRGFLPSADSDLKSSXXSE 234
orfl36ng      AFAGTVYRVFVCLFYIINDGIAHHTAPQRVRYLFAPYRGFLPPASDSDLKSSKYSE 235

```

The complete length ORF136ng nucleotide sequence <SEQ ID 555> is:

```

1 ATGATGAAGC GCGTATAGC CGTCTTCGTC CTGCTCATGC AGAAAAATCCG
51 GATTTTGGGA CAACTGTTCG CGAAATACGT CAATACAGTT CCGGCACATC
101 GGATGCTCTT CCAAATTTTC GGGATGTTCT TTTTCTCATC ACACCGGCAG
151 TACCTGCCCG GGATCGCCGA AATCGATTCC CCAGGCGGTA TCGTGTTCGG
201 TACGCTCCTC TTCGTCATC TGTCGCGCGA TTGCTGTAC GGTAAAGCCG
251 CCGTAGGGGA TCGCGTTGCA CACGACATC CAGTCTGTA GGTAAAGCCG
301 CGGAACGCAG ACGCTTTCG CTGCTTCAG TCGTCTGTA GGTAAAGCCG
351 CTTTGTTCAG CACACCGAAT ATATAAAGAC CCGTCTGTA GGTAAAGCCG
401 ATCCCATAT GTTCGCAAAAT TTCGCGCTCT TCGCGCTCTT GGAAAAAAGG
451 GACTTTGACC ATGCGCAAAAT CCAAGCGCGA AATAATGGCG CGGCGTTCCC
501 AAAAAAGCTC GCGCCAAAGT TATTGGAATG TTTTACGGGC GCGTTTCGCG
551 GCACGGTTTA CCGGTTTCGT TGCCTGTTCT ACATAATAAA TGACGGAATC
601 GCCCATCATA CTGCTCCTCA ACGTGTACGG TATCTGTTTG CACCTTACCG
651 CGGTTTCTTA CCTCGGCAT CCGATTCCGA TTTGAAAAGT TCCAATATT
701 CGGARTAG

```

This encodes a protein having amino acid sequence <SEQ ID 556>:

```

1 MMKRIIAVFV LLMQKIRILG QLLPKIVNTV PAHRMLFQIF GMFFFFFIHRQ
51 YLPGIAEIDS PGGIVFGTLL FRHLSAHCLY GKAAVGDAVA HEHPVADVAN
101 RNANAFALFD IGQSAGFIVQ HTVNIKTVKI NIVDPHMFAN FAVFAVLEKR
151 DFDHGKIQQG NNAAPFKKL APKVFECFTG AFAGTVYRVF CLFYIINDGI
201 AHHTAPQVRV YLFAPYRGFL PPASDSDLK SSKYSE*

```

ORF136ng and ORF136-1 show 93.6% identity in 235 aa overlap:

```

45 orfl36ng      MMKRIIAVFVLLMQKIRILGQLLPKIVNTVPAHRMLFQIFGMFFFFFIHQQYLPGIAEIDS
orfl36-1      MMKRIIAVFVLLMQKIRILGQLLPKIVNTVPAHRMLFQIFGMFFFFFIHQQYLPGIAEIDS
50 orfl36ng      PGGIVFGTLLFRHLSAHCLYGKAAVGDAVAHEHPVADVNRNANAFALFDIGQSAGFIVQ
orfl36-1      PCGIVFGALLFRHLPAHCLYGKAAVGDAVAHEHPVADVNRNANAFALFDIGQFAGFIVQ
55 orfl36ng      HTVNIKTVKINIVDPHMFANFAVFAVLEKRDFDHGKIQQGNNAAPFKKLAPKVFECFTG
orfl36-1      HTVNIKTVKINIVDPHMFANFAVFAVLEKRDFDHGKIQQGNNAAPFKKLAPKIFECFTG
orfl36ng      AFAGTVYRVFVCLFYIINDGIAHHTAPQRVRYLFAPYRGFLPPASDSDLKSSKYSEX
orfl36-1      AFVGTVYRVFVCLFYIINDGIAHHSAPQRVRYLFAPYRGFLPPASDSDLKSSKYSEX

```

Based on the presence of the putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 67

5 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 557>:

```

1 ATGGAATAA TGGTAACGTT TTCAAAATC AGACCGCTTT TGGCAATCGC
51 CGCGCGCGCG TTGCTTGCCG CC.TGCGGAC GCGGGGAAAT AATGCTGTCC
101 GCAAGCCGGT GCAAAACGCC AAACCGCGCG CAGTGGTCGG TTTGGCAGTC
151 GGTGGCGCGG CATCTAAAGG ATTTGCCCAT GTAGGTATTA TTAAGSTTTT
201 GAAGAAATAC GGTATTCCTG TGAAGSTGGT TACCGGACCC TCCGCGAGTT
251 GATTTCGGG CAACCTTTT TGAAGCGGTA TGTGCGCGGA CCGCCTCGAA
301 TTGAAGCCCG AATTTTAGG CAAACCCGAT TTGTCGATTT TAACCTTTTC
351 CACCAATGGG TTATCAAGG GCGCAAGCT GCAAAATTAC ATCAACCGAA
401 AACTCGCGG CATGCAGATT CAGCAGTTTC CCATCAATTT TGCCGCC..

```

15 This corresponds to the amino acid sequence <SEQ ID 558; ORF137>:

```

1 MENMVTFSKI RPLLAIAAAA LLAAXRTAGN NAVRKPVQTA KPAAVVGLAL
51 GGGASKGFAH VGIIKVLKEN GIPVKVVTGT SAGSIVGNLF ASGMSPDRL
101 LEAEILGKTD LVDLTLSTNG FIKGAKLQNY INRKLRGMQI QQFPFIKFAA..

```

Further work revealed the complete nucleotide sequence <SEQ ID 559>:

```

1 ATGGAATAA TGGTAACGTT TTCAAAATC AGACCGCTTT TGGCAATCGC
51 CGCGCGCGCG TTGCTTGCCG CTGCGGCGAC GCGGGGAAAT AATGCTGTCC
101 GCAAGCCGGT GCAAAACGCC AAACCGCGCG CAGTGGTCGG TTTGGCAGTC
151 GGTGGCGCGG CATCTAAAGG ATTTGCCCAT GTAGGTATTA TTAAGSTTTT
201 GAAGAAATAC GGTATTCCTG TGAAGSTGGT TACCGGACCA TCGGCGAGTT
251 GATTTCGGG CAGCCTTTT GCAATCGGTA TGTGCGCGGA CCGCCTCGAA
301 TTGAAGCCCG AATTTTAGG CAAACCCGAT TTGTCGATTT TAACCTTTTC
351 CACCAATGGG TTATCAAGG GCGCAAGCT GCAAAATTAC ATCAACCGAA
401 AAGTCGCGG CAGGCAGATT CAGCAGTTTC CCATCAATTT TGCCGCCGTT
451 GCTACTGATT TTGAACCGG CAAGGCGCTC GCTTTCAATC AGGGGAATGC
501 CGGGCAGGCT GTGCGCGCTT CCGCGCGCAT TCCCAATGTG TTCCAACCGG
551 TTATCATCGG CAGGCATACA TATGTTGACG GCGGCTGTGC GCAGCCGCTG
601 CCGTCAGTGT CCGCGCGCGG GCAGGGGGCG AATTTCTGTG TTAGCGTCGA
651 TATTTCCGCC GTCGCGGGCA AAAACATCAG CCAAGGTTTC TTCTCTTATC
701 TCGATCAGAC GCTGAACGTA ATGAGCTTTT CTGCGTGTGA AATGAGTTG
751 GGGCAGGCGG ATGTGTTTAT CAAACCGCAG GTTTTGATTT TGGTGTGAT
801 CGGCGGATTC GATCAGAAAA AACGCGCATC COGTTTGGT GAGGAGGCG
851 CAGTCGCGG ATTGCCTGAA ATCAACGCA AACTGCGCGG ATACGTTTAT
901 TGA

```

This corresponds to the amino acid sequence <SEQ ID 560; ORF137-1>:

```

1 MENMVTFSKI RPLLAIAAAA LLAACGTAGN NAVRKPVQTA KPAAVVGLAL
51 GGGASKGFAH VGIIKVLKEN GIPVKVVTGT SAGSIVGSLF ASGMSPDRL
101 LEAEILGKTD LVDLTLSTSG FIKGKRLQNY INRKVGGRQI QQFPFIKFAV
151 ATDFETGKAV AFNGNAGQCA VRASALPNV QQPVIIIRHT VVDGGLSQPV
201 PVSAARRQGA NFVIADVISA RPKGNISQGF FSLDQTLNV MSVSLQNEL
251 QQADVVIKPO VLDLGAUVGF DQKKRAIRLG EEAARAALPE TKRKLAAARY
301 *

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF137 shows 93.3% identity over a 149aa overlap with an ORF (ORF137a) from strain A of *N.*

meningitidis:

		10	20	30	40	50	60
	orf137.pep	MENMVTFSKIRP	LLAIAAAAL	LAAXRTAG	NNAVRKPVQ	TAKPAAVVGL	ALGGGASKGFAH
5	orf137a	MENMVTFSKIRP	LLAIAAAAL	LAACGTAG	NNAARKPVQ	TAKPAAVVGL	ALGGGASKGFAH
		10	20	30	40	50	60
	orf137.pep	VGIIKVLKENG	IPVKVVTG	TSAGSIVGN	LFASGMSPD	RLEAEILGK	TDLVDLTLS
10	orf137a	VGIIKVLKENG	IPVKVVTG	TSAGSIVGN	LFASGMSPD	RLEAEILGK	TDLVDLTLS
		70	80	90	100	110	120
	orf137.pep	FIKGEKLQNY	INRKLGRM	QIQQFPIK	FAA		
15	orf137a	FIKGEKLQNY	INRKLGRM	QIQQFPIK	FAA		
		130	140	149			
	orf137.pep	FIKGEKLQNY	INRKLGRM	QIQQFPIK	FAA		
	orf137a	FIKGEKLQNY	INRKLGRM	QIQQFPIK	FAA		
		130	140	150	160	170	180

The complete length ORF137a nucleotide sequence <SEQ ID 561> is:

20	1	ATGGAAATA	TGGTAACGTT	TTCAAAATC	AGACCGCTTT	TGGCAATCGC
	51	CGCGCCCGC	GTGCTTGCCG	CCTGCGGCAC	GGCGGGAAT	AATGCTGCCC
	101	GCAAGCCGG	GCAACCGCC	AAACCCGCC	CAGTGGTCGG	TTTGGCACTC
	151	GGTGGCGCG	CATCTAAAGG	ATTTGCCAT	TAGGTATTA	TAAAGTTT
	201	GAAAGAAAC	GGTATTCCTG	TGAAGTGGT	TACCGGCACA	TCGGCAGGTT
25	251	CGATAGTCG	GAGCCTTTT	GCATCGGTA	TGTGCGCCGA	CGCGCTCGAA
	301	TTGGAAGCG	AAATTTTAGG	TAAAACCGAT	TTGGTCGATT	TAACTTTCG
	351	CACCACTGT	TTTATCAAG	CGCAAAAGT	GCATAATAC	ATCAACCGAA
	401	AAGTCGGCG	CAGGCGGATT	CAGCAGTTC	CCATCAAAAT	TGCGCGCGTT
	451	GCTACTGAT	TTGAACCGG	CAGGCGCGT	GCCTTCAATC	AAGGGAATGC
30	501	CGGGCAGGT	GTGCGCGCT	CCGCGCCAT	TCCCAATGTG	TTCCAACCGC
	551	TTATCATCG	CAGGCATACA	TATGTTGAC	GCGGTCGTGC	GCAGCCCGTG
	601	CCGCTCAGT	CGCGCGCGG	GCAGNNNNN	NATNTCGTGA	TTCGCGTGA
	651	TATTTCCGC	CGTCCGAGCA	AAAACATCAG	CCAGGCTTTC	TTCTCTTATC
	701	TGATCAGAC	GCTGAACGTA	ATGAGCGTTI	CCGCGTTCGA	AAATGAGTGT
35	751	GGGCAGGCG	ATGTGCTTAT	CAAACCGCAG	CTTTTGGATT	TGGTGCAT
	801	CGCGCGATT	GATCAGAAAA	AACCGCCAT	CCGTTGGGT	GAGGAGCGAG
	851	CAGCTGCCG	ATTGCTTGAA	ATCAACGCA	AACTGGCGCG	ATACCGTTAT
	901	TGA				

This encodes a protein having amino acid sequence <SEQ ID 562>:

40	1	MENMVTFSKI	RPLAIAAAA	LAAAGTAGN	NAARKPVQTA	KPAAVVGLAL
	51	GGGASKGFAH	VGIIKVLKEN	GIPVKVVTG	SAGSIVGSLF	ASGMSPDRL
	101	LEAEILGKTD	LVDLTLSG	FIKGEKLQNY	INRKGGRRI	QQFPIKFAAV
	151	ATDFETGKAV	AFNQNGAGQA	VRASAAIPNV	FQPVIIIGRHT	YVDGGLSQPV
	201	PVSAARRXXX	XXVIAVDISA	RPSKNISQGF	FSYLDQTLNV	MSVSALQNEL
45	251	GQADVVIKQP	VLDLGVGGF	DQKKRAIRLG	EEAARAALPE	IKRKLAARYY
	301	*				

ORF137a and ORF137-1 show 97.3% identity in 300 aa overlap:

	orf137a.pep	MENMVTFSKIRP	LLAIAAAAL	LAACGTAGN	NAARKPVQTA	KPAAVVGLAL	GGGASKGFAH
	orf137-1	MENMVTFSKIRP	LLAIAAAAL	LAACGTAGN	NAARKPVQTA	KPAAVVGLAL	GGGASKGFAH
50	orf137a.pep	VGIIKVLKENG	IPVKVVTG	TSAGSIVGN	LFASGMSPD	RLEAEILGK	TDLVDLTLS
	orf137-1	VGIIKVLKENG	IPVKVVTG	TSAGSIVGN	LFASGMSPD	RLEAEILGK	TDLVDLTLS
55	orf137a.pep	FIKGEKLQNY	INRKGGRRI	QQFPIKFAA	VATDFETG	KAVAFNQNG	AGQAVRASAAIPNV
	orf137-1	FIKGEKLQNY	INRKGGRRI	QQFPIKFAA	VATDFETG	KAVAFNQNG	AGQAVRASAAIPNV
60	orf137a.pep	FQPVIIIGRHT	YVDGGLSQ	PPVPSAARR	XXVIAVDIS	ARPSKNISQ	GFESYLDQTLNV
	orf137-1	FQPVIIIGRHT	YVDGGLSQ	PPVPSAARR	XXVIAVDIS	ARPSKNISQ	GFESYLDQTLNV
	orf137a.pep	MSVSALQNEL	GQADVVIKQP	VLDLGVGGF	DQKKRAIRL	GEEAARAAL	PEIKRKLAARYY

orf137-1 MSVSALQNELGQADVVIKPVLDLGA VGGFDQKKRAIRLGEAARAALFEIKRKLAAARY

Homology with a predicted ORF from *N.gonorrhoeae*

- 5 ORF137 shows 89.9% identity over a 149aa overlap with a predicted ORF (ORF137ng) from *N.gonorrhoeae*:

orf137.pep	MENMVTFSKIRPLLAIAAAALLAAXRTAGNNAVRKPVQTA	KPAAVVVALG	GGGASKGFAH	60
orf137ng	MENMVTFSKIRSLAIAAAALLAACGTAGNNAARKPVQTA	KPAAVVVALA	GGGASKGFAH	60
orf137.pep	VGIIKVLKENGIPVKVVTGTSAGSIVGNLFASGMS	PDRLLEAEIL	LGKTDLVDLTLSTNG	120
orf137ng	IGIVKVLKENGIPVKVVTGTSAGSIVGSLASGMS	PDRLLEAEIL	LGKTDLVDLTLSTSG	120
orf137.pep	FIKGAQLQNYINRKLGMQIQQFPIKFAA			149
orf137ng	FIKGEKLQNYINRKVGGRQIQFPIKFAAVATDFETG	KAVAFNQGNAGQAV	VRASAAIPNV	180

The complete length ORF137ng nucleotide sequence <SEQ ID 563> is:

1 ATGGAATAAATGTTACGTTTCAAAATCAGATCATTTTGGCAATCGC
 20 51 CGCGCGCGCGTGTCTGCGCGCTGCGGTACGGCGGAAACAAATCGCGCC
 101 GCAGCGCGGTGCAACCGCCAAACCGCGCGCAGTGGTCGCTTTGGCACTC
 151 GGTGGCGCGCATCTAAAGATTTCGCCATATAGGAATGTTAAGGTTT
 201 GAAAGAAACGGTATCTCTGTGAAGTGGTTACCGGCACATCGCGAGTT
 25 251 CGATAGTCGCGAGCCTTTGCGATCGGGTATGTCGCGCGAATCGCGTAA
 301 TTGGAAGCGAGATTTTAGTAAACCGGTTAGTCGATTAACTTGTCT
 351 CACCACTGGTTTATCAAAAGCGAAGAGCTGCAAAATTACATCAACGAA
 401 AAGTCGCGCGCAGCGAGATTGAGCAGTTTCCATCAAAATTTGCGCGGTT
 451 GCCACTGATTGTGAACCGCGAAGCGCGCTGCTTCAATCAAGGAATCG
 501 CGGCGAGCGGTTGCTGCTCTCGCGCGCATTCGCAATGTTCCAGCGAG
 551 TCATCATCGCGAGGACACAAATAGTTTACGCGGCTCTGCGCGCGT
 601 CCGTCAGCTGCGCTCGCGCGCAGTTCGCGCAATTCGCGCTGCA
 651 TATTTCGCGCAGTTCGCGCAAAATGTCGCTCAAGGTTTCTCTTATC
 701 TGATCAGCAGCTGAACGTGATGAGCGTTTCCGTTGTGCAAAACGAGTT
 751 gggcAGGCGGATGTCGTTATCAAACGCGaggtTTTGGATTGGGTGCA
 801 CGCGCGGATTCGATCAGAAAAAGCGCGCCATCCGTTTGGCGAGGAGGAG
 851 CACGTGCGCGCATTCGCTGAAATCAACGCAACATGCGCGCATACCGTTAT
 901 TGA

This encodes a protein having amino acid sequence <SEQ ID 564>:

1 MENMVTFSKIRSLAIAAAALLAACGTAGNNAARKPVQTA
 40 51 GGGASKGFAHIGIVKVLKENGIPVKVVTGTSAGSIVGSLASGMS
 101 LEAEILGKTDLVDLTLSTSGFIKGEKLQNYINRKVGGRQIQFPIKFAAV
 151 ATDFETGKAVAFNQGNAGQAVRASAAIPNVFQVPIGRHKYVDGGLSQV
 201 PVSAARRQGANFVIAVDISARPSKNVGGFFSYLDQTLINVMSVVLQNEL
 251 QGADVVIKPVLDLGA VGGFDQKKRAIRLGEAARAALFEIKRKLAAARY
 301 *

ORF137ng and ORF137-1 show 96.0% identity in 300 aa overlap:

orf137ng	MENMVTFSKIRSLAIAAAALLAACGTAGNNAARKPVQTA	KPAAVVVALA	GGGASKGFAH
orf137-1	MENMVTFSKIRPLLAIAAAALLAACGTAGNNAVRKPVQTA	KPAAVVVALG	GGGASKGFAH
orf137ng	IGIVKVLKENGIPVKVVTGTSAGSIVGSLASGMS	PDRLLEAEIL	LGKTDLVDLTLSTSG
orf137-1	VGIIKVLKENGIPVKVVTGTSAGSIVGSLASGMS	PDRLLEAEIL	LGKTDLVDLTLSTSG
orf137ng	FIKGEKLQNYINRKVGGRQIQFPIKFAAVATDFETG	KAVAFNQGNAGQAV	VRASAAIPNV
orf137-1	FIKGEKLQNYINRKVGGRQIQFPIKFAAVATDFETG	KAVAFNQGNAGQAV	VRASAAIPNV
orf137ng	FQVPIIGRHKYVDGGLSQVFPVSAARRQGANFVIAVD	ISARPSKNVGGFFSYLDQTLINV	


```

orf137-1      FQPVIIGRHTYVDGGLSQPVFVSAARRQGANFVIAVDISARPGKNISQGGFFSYLDQTLNV
orf137ng      MSVSVLQNELGQADVVIKQVLDLGAUGGFDQKKRAIRLGEEAARAALPEIKRKLAAARY
1             ||||:|||||
orf137        MSVSALQNELGQADVVIKQVLDLGAUGGFDQKKRAIRLGEEAARAALPEIKRKLAAARY

```

Based on the presence of a predicted prokaryotic membrane lipoprotein lipid attachment site (underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 68

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 565>:

```

1  ATGTTTCGTT  TACAATTCAG  GCTGTTTCCC  CCTTTGCGAA  CCGCCATGCA
51 CATCCTGTGTG ACCGCCCTGC  TCAATGCCT  CTCCTGcTG  CCGCTTTCCT
101 GTCTGCACAC  GCTGGGAAC  CGGCTCGGAC  ATCTGGCGTT  TTACCTTTTA
151 AAGGAAGACC  GCGCGCGCAT  CGTCGCGcAT  ATCGCGCAGG  CGGGTTTGAA
201 CCCCGACCCC  AAAACGGTCA  AAGCGTTTT  TCGCGAAACG  GCAAAGGCG
251 GTTTGGAAct  TGCCCCCGCG  TTTTTCAGAA  AACCGGAAGA  CATAGAACA
351 ATGTTCAAAG  CGGTACACGG  CTGGGAACAT  GTGCGACAGG  CTTTGGACAA

```

This corresponds to the amino acid sequence <SEQ ID 566; ORF138>:

```

1  MFRLQRLFP  PLRTAMHILL  TALLKCLSL  PLSCSLTIGN  RLGLHAFYLL
51 KEDRARIVAX MRQAGLNPD  KTVKAVFAET  AKGGLELAPA  FFRKPEDTET
101 MFKAVHGW  VQQLDXHEG  LLF

```

Further work revealed the complete nucleotide sequence <SEQ ID 567>:

```

1  ATGTTTCGTT  TACAATTCAG  GCTGTTTCCC  CCTTTGCGAA  CCGCCATGCA
51 CATCCTGTGTG ACCGCCCTGC  TCAATGCCT  CTCCTGcTG  CCGCTTTCCT
101 GTCTGCACAC  GCTGGGAAC  CGGCTCGGAC  ATCTGGCGTT  TTACCTTTTA
151 AAGGAAGACC  GCGCGCGCAT  CGTCGCGcAT  ATCGCGCAGG  CGGGTTTGAA
201 CCCCGACCCC  AAAACGGTCA  AAGCGTTTT  TCGCGAAACG  GCAAAGGCG
301 GTTTGGAAct  TGCCCCCGCG  TTTTTCAGAA  AACCGGAAGA  CATAGAACA
351 ATGTTCAAAG  CGGTACACGG  CTGGGAACAT  GTGCGACAGG  CTTTGGACAA
351 ACACGAAGGG  CTGCTATTCA  TCAACGCCGA  CATCGGCAGC  TACGATTGCG
401 CGGACGCGTA  CATCAGCCAG  CAGCTTCCGT  TCCCGCTGAC  CGCATGTGAC
451 AAACCGCGCA  AATCAAAGC  GATAGACAA  ATCATGCAGG  CGGGCAGGGT
501 TCGCGGCAAA  GGAARAACCG  CGCCTACAG  CATACAGGG  GTCAACAAGA
551 TCATCAAAGC  CTTGCGTTCC  GCGCAAGCAA  CCATCGTCTT  CGCCGACAC
601 GTCCCTCCCT  CTCAGAAGG  CGGGGAAGGC  GTATGGGTGG  ATTTCTTCGG
651 CAAACCTGCC  TATACCATGA  CGCTGCGCG  AAAATTGGCA  CACGTCAAAG
701 GCGTGAAAC  CTTGTTTTTC  TGCTGCGAAC  GCCTGCCTGG  CGGACAAGGT
751 TTCGATTTC  ACATCCGCC  CGTCAAGGC  GAATTGAACG  CGGACAAGGT
801 CCATGATGCC  GCGCTGTTC  ACCGCAATGC  CGAATTATTG  ATACGCCGTT
851 TTCCGACGCA  GTATCTGTT  ATGTACAAC  GCTACAARAT  GCGGTAA

```

This corresponds to the amino acid sequence <SEQ ID 568; ORF138-1>:

```

1  MFRLQRLFP  PLRTAMHILL  TALLKCLSL  PLSCSLTIGN  RLGLHAFYLL
51 KEDRARIVAN MRQAGLNPD  KTVKAVFAET  AKGGLELAPA  FFRKPEDTET
101 MFKAVHGW  VQQLDXHEG  LLFIPTHIG  YDGGRIYISQ  LLFFPLITAMY
151 KPKKALDK  IMQAGRVK  GKAPTSTIQ  VKGIKALKRS  GEATIVLPDH
201 VPSFQEGEG  VVWDFEKRA  YMTLAAKLA  HVKGVKTLFF  CCRSLPGGCG
251 FDLHIRPVQ  ELNGDKAHD  AVFNNAEY  IRRFPTQYLE  MYNRKPM*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF138 shows 99.2% identity over a 123aa overlap with an ORF (ORF138a) from strain A of *N.*

meningitidis:

		10	20	30	40	50	60
5	orf138.pep	MFRQLQRLFPPLRTAMHILLTALLKCLSLPLSCLHTLGNRLGHLAFYLLKEDRARIVAX					
	orf138a	MFRQLQRLFPPLRTAMHILLTALLKCLSLPLSCLHTLGNRLGHLAFYLLKEDRARIVAN					
		10	20	30	40	50	60
10	orf138.pep	MRQAGLNPDPKTVKAVFAETAKGGLELAPAFFRKPEDIETMFKAVHGWHEHVQALDKHEG					
	orf138a	MRQAGLNPDPKTVKAVFAETAKGGLELAPAFFRKPEDIETMFKAVHGWHEHVQALDKHEG					
		70	80	90	100	110	120
15	orf138.pep	LLF					
	orf138a	LLFITPHIGSYDLGGYISQQLPFLPTAMYKPPKIKAIKIMQAGVRVKGKGTAPTSTIQG					
		130	140	150	160	170	180

The complete length ORF138a nucleotide sequence <SEQ ID 569> is:

```

1  ATGTTCTGTT  TACAAATCAG  GCTGTTTCCC  CTTTTCGAA  CCGCCATGCA
51  CATCCTGTGT  ACCGCCCTGC  TCAAATGCCT  CTCCTCGCTG  CCGCTTCCCT
101  GTCTGCACAC  GCTGGGAAC  CGCGTCGGAC  ATCTGGCGTT  TTACCTTTTA
151  AAGGAAGACC  GCGCGCGCAT  CGTCGCCAAT  ATCGCTCAGG  CAGGCATGAA
201  TCCGACACCC  AAAACGGTCA  AAGCCGTTTT  TCGCGAAGCG  GCAAAAGCGG
251  GTTTGGAAC  TGCCCGCCGG  TTTTTCAGAA  AACCGGAAGA  CATAGAACA
301  ATGTTCAAAG  CGGTACACGG  CTGGAAACAT  GTGCGACAGG  CTTTGGACAA
351  ACACGAGGG  CTGCTATTCA  TCACGCCCA  CATCGCGTAC  TACGATTGCG
401  CGGACGCTA  CATCACCGG  CAGCTCGGT  TCCGCTTAC  CGCGATTGAC
451  AACCGCCGA  AATCCAGAC  GATAGACAAA  ATCATCGAGG  CGCGAGGCT
501  TCGCGCCAA  GGAACAAAC  CGCCTACAC  CATACAAGG  GTCAACAAA
551  TCATCAAAGC  CCGTGGTTCG  GCGGAAGCAA  CATCGCTCCT  GCCCGACCAC
601  GTCCCTCCC  CTCAGAAGG  CGGGAAGCG  GTATGGGTGG  ATTTCTTCGG
651  CAAACCTGCC  TATACCATGA  CGCTGGCGCG  AAAATTGGCA  CACGTCAAAG
701  GCGTGAAAC  CCTGTTTTTC  TGCTGCGAAC  GCCTGCTGG  CGGACAAGTG
751  TTCGATTTC  ACATCCGCCC  CGTCAAAGG  GAATTGAAC  GCGACAAGCG
801  CCATGATGCC  GCCGTGTTCA  ACGCAATGC  CGAATATTGG  ATACCGCGGT
851  TTCGACGCA  GTATCTGTTT  ATGTACAACC  GCTACAAAT  GCCGTAA

```

This encodes a protein having amino acid sequence <SEQ ID 570>:

```

1  MFRQLQRLFP  PLRTAMHILL  TALLKCLSL  LPLSCLHTL  GNRLGHLAF  YLLKEDRA  RIVAN
51  KEDRARIVAN  MRQAGLNPD  KTVKAVFAE  AKGGLELAP  FFRKPEDIE  T
101  MFKAVHGWHE  VQALDKHEG  LLFITPHIG  SYDLGGYIS  QQLPFLPTA  MY
151  KPPKIKAIK  IMQAGVRGK  GKTAPTST  IQG VKQIIK  ALRS GEATVLP  DH
201  VPSPEGGEG  VVWDFGKPA  YMTLAAKLA  HVKGVKTL  FF CCERLPG  GQG
251  FDLHIRPVQ  ELNGDKAHD  AVFNRNAE  YW IRRFPTQ  YLF MYNRYK  MP*

```

ORF138a and ORF138-1 show 99.7% identity over a 298aa overlap:

	orf138a.pep	MFRQLQRLFPPLRTAMHILLTALLKCLSLPLSCLHTLGNRLGHLAFYLLKEDRARIVAN	
	orf138-1	MFRQLQRLFPPLRTAMHILLTALLKCLSLPLSCLHTLGNRLGHLAFYLLKEDRARIVAN	
	orf138a.pep	MRQAGMNPDPKTVKAVFAETAKGGLELAPAFFRKPEDIETMFKAVHGWHEHVQALDKHEG	
	orf138-1	MRQAGLNPDPKTVKAVFAETAKGGLELAPAFFRKPEDIETMFKAVHGWHEHVQALDKHEG	
	orf138a.pep	LLFITPHIGSYDLGGYISQQLPFLPTAMYKPPKIKAIKIMQAGVRVKGKGTAPTSTIQG	
	orf138-1	LLFITPHIGSYDLGGYISQQLPFLPTAMYKPPKIKAIKIMQAGVRVKGKGTAPTSTIQG	
	orf138a.pep	VKQIIKALRSGEATVLPDHSVPSPEGGEGVWVDFGKPAYMTLAALAHVKGKVTLLFF	

```

      |||
orf138-1  VQKI I KALRSGEATIVL PDHVSPQEGGEGVWVDFGKPAYTMTLA AKLAHVKGWKT LFF
      |||
5  orf138a.pep  CCE RLPGGQGF DLH I R F V Q G L N G D K A H D A A V F N R N A E Y W I R R F P T Q Y L F M Y N R Y K M P
      |||
orf138-1  CCE RLPGGQGF DLH I R F V Q G L N G D K A H D A A V F N R N A E Y W I R R F P T Q Y L F M Y N R Y K M P
      |||

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF138 shows 94.3% identity over a 123aa overlap with a predicted ORF (ORF138ng) from

10 *N.gonorrhoeae*:

```

orf138.pep  MFR LQFR LFP LRTAMH ILL TALLKCLSL LPLSCLHT LGNRLGH LAFYLLKEDRARI VAX 60
      |||
orf138ng    MFR LQFR LFP LRTAMH ILL TALLKCLSL LPLSCLHT LGNRLGH LAFYLLKEDRARI VAN 60

15 orf138.pep  MRQAG LNPDPK TVKAVFAETAKG GLELAPAFFRKPEDIETM FKA VHGWEHVQ QALDKHEG 120
      |||
orf138ng    MRQAG LNPDTQ TVKAVFAETAKG GLELAPAFFRKPEDIETM FKA VHGWEHVQ QALDKGEG 120

20 orf138.pep  LLF 123
      |||
orf138ng    LLFTTPHIGSYDLGGRIYSQQLPFHLTAM YKPPKIKAI DKIMQAGRVRGKGKTAPTGIQG 180

```

The complete length ORF138ng nucleotide sequence <SEQ ID 571> is:

```

1  ATGTTTCGTT TACAATTCAG GCTGTTTCCC CTTTGCAGAA CCGCATGCA
25 51  CATCCTGTGTG ACGGCCCTGC TCAAAATGCCT CTCCTGCTGT TCGCTTTCCCT
101 GTCTGCACAC GCTGGGAAAC CGGCTCGGAC ATCTGGCGTT TTACCTTTTA
151 AAGGAAGACC GCGCGCGCAT CGTCGCCAAT ATGCGGCAGG CGGGTTTGAA
201 CCCCAGACAC CAGACGGTCA AAGCGTTTT TCGGGAACGG GCAGAAATGG
251 GTTTGGAACT TGCCCCCGCG TTTTCAAAA AACCGAAGA CATCGAACA
301 ATGTTCAAA GCGTACACG CTTGGACAG CTTGACACA CTTTGGACA
351 GCGCGAAGC CTGCTTTCA TCACGCCGC CATCGCAGC TACGATTGG
401 CCGGAGCGTA CATCAGCGAC CAGCTCCGT TCCACCTGAC CGCCATGTAC
451 AAGCGCGCA AATCAAAAG GATAGACAAA ATCATGCGAG CGGGCAGGTT
51 GCGCGGCAGG GCAAAAACcg cgcacacgg catACAAGSG GTCAACACAA
35 551 tcatcaAGGC CCTGCGCGCG GCGGAGGCAA CcATcATCCT GCCCGACCA
601 GTCCCTTCTC CGCAGGAagg cggCGGCGTG TGGCGGATT TTTTCGGCAA
651 ACCTGCATac acCATGACAC TGGCGGCAA ATTGGCACAC GTCAAGGCG
701 TGAAAACCC TTTTCTCTG TCGGAACGCC TGCCCGACGG ACAAGGCTT
751 GTGTTGCACA TCCGCCCGT CCAAGGGGAA TTGAACGCCA ACAAGGCCA
801 CGATGCCGCC GTGTTCAACC GCAATACCGA ATATTGGATA CGCCGTTTC
40 851 CGACGCGATA TCTGTTTATG TACAAACCGT ATAAACGCC GTAA

```

This encodes a protein having amino acid sequence <SEQ ID 572>:

```

1  MFR LQFR LFP LRTAMH ILL TALLKCLSL LPLSCLHT LGNRLGH LAFYLLKEDRARI VAN
51  KEDRARI VAN MRQAG LNPDP TVKAVFAET AKCGLELAP FFKKPEDIBT
45 101 MFKAVHGWEH VQALDKGEG LLFTTPHIGS YDLGGRYSQ QLPHLTAMY
151 KPPKIKAI DKIMQAGRVRGK GTAPTGIQG VKQIKALRA GEARTILPH
201 VPSQEGG V WADPFKGPAY TMTLA AKLAHV KGWKT LFFC CERLPDQCGF
251 VLH I R F V Q G L N G D K A H D A A V F N R N E Y W I R R F P T Q Y L F M Y N R Y K T P *

```

ORF138ng and ORF138-1 show 94.3% identity over 299aa overlap:

```

50 orf138-1.pep  MFR LQFR LFP LRTAMH ILL TALLKCLSL LPLSCLHT LGNRLGH LAFYLLKEDRARI VAN
      |||
orf138ng    MFR LQFR LFP LRTAMH ILL TALLKCLSL LPLSCLHT LGNRLGH LAFYLLKEDRARI VAN

55 orf138-1.pep  MRQAG LNPDPK TVKAVFAETAKG GLELAPAFFRKPEDIETM FKA VHGWEHVQ QALDKHEG
      |||
orf138ng    MRQAG LNPDTQ TVKAVFAETAKG GLELAPAFFRKPEDIETM FKA VHGWEHVQ QALDKGEG

orf138-1.pep  LLFTTPHIGSYDLGGRIYSQQLPFHLTAM YKPPKIKAI DKIMQAGRVRGKGKTAPTSTIQ
80 orf138ng    LLFTTPHIGSYDLGGRIYSQQLPFHLTAM YKPPKIKAI DKIMQAGRVRGKGKTAPTGIQG

```

[illegible]

In addition, ORF138ng is homologous to htrB protein from *Pseudomonas fluorescens*:

```
gnl|FID|e334283 (Y14568) htrB [Pseudomonas fluorescens] Length = 253
Score = 80.8 bits (1468), Expect = 9e-15
Identities = 49/151 (32%), Positives = 79/151 (51%), Gaps = 6/151 (3%)

Query: 101 MFAVHGWEHGWEOALQDLKDEGLLPIFPHGISYD-LGGRYISQQLPFLHTAMYKPKKIKAD 159
      + + V G E + + + A L G + + + I T H + + + L Y S Q P + + P P K + K A + D
Sbjct: 94 LVREVEGLLEVLKEALASGKGVGITSHLNGVFNVLNHFYCSQCKPI--IIFYRPFKLKAVD 150

Query: 160 KIMQAGRVGRGKGTAPTGI+QGVQKIIKALRAGEATIIIPDVPSPQEGGVWADFFGKPA 219
      + + + + + R V + K + A + + + + + I K + R + E D F E E F F A
Sbjct: 151 ELLRQRVRVLGMNKVAASTKEGILSVIKERGGQGVIGPAD--PEFASAGIFVFPEATQA 208

Query: 220 TMTLAARLAAHVKGKVTLEFICCLRLPDGQG 250
      Y + + + + + F +
Sbjct: 209 ITSKFVNPNMLGAGKGVGLFHALRLPDGSG 230
```

Based on this analysis, including the presence of a putative transmembrane domain in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF138-1 (57kDa) was cloned in the pGex vectors and expressed in *E.coli*, as described above.

The products of protein expression and purification were analyzed by SDS-PAGE. Figure 14A shows the results of affinity purification of the GST-fusion protein. Purified GST-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result) and FACS analysis (Figure 14B). These experiments confirm that ORF138-1 is a surface-exposed protein, and that it is a useful immunogen.

Example 69

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 573>:

```

1      .CGCGTGGTGGG      CGCGCGGAATC      GTGGCGGTGTG      TTAATGTGAAA      GTGAACCGTG
51     GCATCGCGGTG      TGGAAATACTT      TGCGCCTTGTG      CGGGCGCGCGG      GTGTATCGGG
101    CACGCGGTTT      TGGTGTGTGTG      TATCGCGCGCG      CGGCGCGCGGG      GTGTGCGGTG
151    ATGCGCGCGG      TGATGTTTIA      GCGGCTTAATG      GCGGCGCGCGG      TTGTGTGCTG
201    GCGCGCGGTG      ATGCGCGGTG      ATGCGCGGTG      ATGCGCGGTG      ATGCGCGGTG
251    TGCTGCGCAAT      TGATCGCGCTG      CGCGCGGATC      CGTTTGTGGC      AARAGATGTT
301    TTATCAGCTT      GGGATGCACT      CGCGCGCGAT      TACGCGCAAG      CGGCGCGCGG
351    TTTGGTGCA      AACGCGCTTC      AGACGCGGAT      CGCGATCAGT      TTTCCCTCTG
401    TGAAGAACCGG      TGTGCGCGGTG      GGTGCTGACTT      TGCGGCGCGG      ACCTCGCGTG
451    GCGCAAGATTG      CGCGCGCAAT      GTTTTGTGCTG      TCGTCCGAAT      GCGCACGCGT
501    GACGATGCTG      ATGTCGCGGT      GTTTTGGGACG      CGCGGCTGAG      GATAATATCG
551    CGCGCGGCAAT      GGTGCTG..

```

This corresponds to the amino acid sequence <SEQ ID 574; ORF139>:

1 ..AWSAGESWRV LMESETWHAV WNTLRFSAAA VYAAAVLGUV YAAPARRSAW
51 MRGLMFXPFM VSPVCVSAGV LLLYPQWTAS LPLLLAMYAL LAYPFVAKDV

101 LSAWDALFPD YGRAAAGLGA NGFQTACRIT FELLKPALRR GLTLAAATCV
 151 GEFAATLFLS RPEWQTLITL TYAYLGRAGE DNYARAMVL..

Further work revealed the complete nucleotide sequence <SEQ ID 575>:

1 ATGGATGGAAC GCGGTGGGGT GGTATGGGGT GCTTTGGGCC TGCTGCCTTC
 51 GCGTTTTTTTG CGGTAATGG CTGTTGGGCC TTGTGGGGCG GTGGCGGGGT
 101 AGACGGGTTT GCGGTGGCGC CCGGTGCTGT CGGATGCCTA TATGCTCAA
 151 CGTTTGGCGT GGACGGTATT TCAGGCAGCG GCAACCTGTG TCGTGGTGGT
 201 CCGTTTGGGC GTGCCGTGCG CGTGGGTGCT GCGCGGGCTG GCGTTTCCGG
 251 GCGCGGCTTT GGTGCTGGCG CTGCTGATGC TGCTTTTGTG GATGCCACAG
 301 TTGGTGGCGG GGTGGGCGGT GCTGGCCCTG TTCGGGGCGG ACGGGCTGTT
 351 GTGGCGCGCG AGGCAGGATA CGCCGTATCT GTTGTGTAC GGCAATGTGT
 401 TTTTCAACCT TCCTGTGTGT GTACAGGCGG CGTATCAGGG GTTTGTGCAA
 451 GTGCGTGGCG CACGGCTTCA GACGGCACGG ACGTGGGCGG CGGGGGCGGTG
 501 GCGGCGGTTT TGGGACATTG AATGCCCCGT TTTGCGCCCG TGGCTTGGCG
 551 GCGGCGGTGT CTTGTCTTTT CTGTATTGTT TTCCGGGTTT CGGGCTGGCG
 601 CTGCTGCTGG GCGGCGAGCG TTATGCCACG GTCCAAATGG AAATTTACCA
 651 GTTGGTCATG TTGCAACTCG ATATGGCGGT TGCTTCGGTG CTGTTGIGSG
 701 TGGTGTGGG GGTAAACGGCG GCGGCGAGGT TGCTGTATGC GTGGTTCGCG
 751 AGGCGCGCGG TTTCCGATAA GCGCGTTTCC CTTGTGATGC CTTGCGCCCG
 801 GCACTGCGCT GGGGAATATG TGCTGGGCGG GTTGGGCGG CGCGTGTGTT
 851 CTGCTGCTGG CCGTCTTCTT TTGTTGCGAA TTGTTGTGAA ACGCTGGTGC
 901 GCGGCGGAAT CGTGGCGGTG GTTAATGGAA AGTGAAGCGT GCGAGCGGCT
 951 GTGGAATACT TTGCGCTTCT CCGCGCGCGG GGTGTATGCG GCGCGGGTTT
 1001 TGGGTGTGGT GTATGCGCGG GCGGCGCGCG GGTGCGCGTG GATGCGCGGG
 1051 CTGATGTTTT TGCCGTTTAT GGTGCGCGCG GTTGTGTGTT CCGCGGGCGT
 1101 GCTGCTGCTT TATCGCGAGT GGACGGCTTC GTGCGCGTTG CTGCTGGCGA
 1151 TGTATGCGCT GCTGCGGTAT CCGTTTGTGG CAAAGATGTT TTTATCAGCC
 1201 TGGGATGCAC TGCCCGCGGA TTACGCGCAG GCGCGCGCGG GTTGGGTGCG
 1251 AAACGGCTTT CAGACGGCAT GCGCGCATAC GTTCCCGCTC TTGAACACGG
 1301 CGTGTGCGCG CGGTCTGACT TTGCGCGCGG CAACCTCGGT GGGCGAATTT
 1351 GCGGCGACAT TGTTCCTGTC GCGTCCGGAA TGGCAGACGC TGAAGACTTT
 1401 GATTATGCC TATTGGGAC GCGCGGTGTA GGATAATTAC GCGCGCGCGA
 1451 TGGTGTGCAC ATTGCTGTTG GCGGCGTTGG CGCTGGGTAT TTCTGCTGTG
 1501 TTGGACGCGG CGGAAGCGG AAAACAGACG GAAACGTTAT AA

This corresponds to the amino acid sequence <SEQ ID 576; ORF139-1>:

1 MDGRRWVVG AFALLPSAFL AVMVVAFLWA VAAVDGLAWR AVLSDAYMLK
 51 RLAWTVFQAA ATCVLVLELG VPVAVLRLR APGRALVLR LLMLEPVMPT
 101 LVAGCVGLAL PGADGLLWRG RQDTFYLLY GNVPFNLVPL VRAAYQCGVFL
 151 VFAARLQAR TLGAGAWRRF WDIEMPVLR WLAGVCLVF LYCFSGFGFLA
 201 LLLGGSRVAT VEVEIYQLVM FELDMVASV LWLVLGVT AAGLLYAWFG
 251 RRAVSDKA VS PVMSPSQSV GEYVLLAFAA AVLSCCLFP LLAIIVKANS
 301 AGESWRVIME SETWQAVWNT LRFSAAVYA AAVLGUVYAA AARRSAAWRRG
 351 LMLPLPMVSP VCVSAGVLLL YEQWTASLEL LLAMAYLLAY PFVAKDVLSA
 401 WDALPFDYGR AAAGLGANGF QACRITTFEL LKPALRRGLT LAATCVGEFF
 451 AATLFLSRPE WQTLTLLIYA YLGRAGEDNY ARAMVTL LL AAFALGIFLL
 501 LDGEGGKQT ETL*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF139 shows 94.7% identity over a 189aa overlap with an ORF (ORF139a) from strain A of *N.*

meningitidis:

				10	20	30
orf139.pep				AWSAGESWRVIMSETHAVWNTLRFSAAA		
orf139a	QSVGEYVLLAFAAAVXSVCCFLXLLAIIVKANS	AGESWRVIMSETHAVWNTLRFSAAA				
	270	280	290	300	310	320
	40	50	60	70	80	90
orf139.pep	VYAAAVLGVVYAAPARRSAAWRRGLMFXPVMSPV	CVSAGVLLYEQWTASLELPLLLAMAYL				

orf139a YVAAAVLGVVYAAAAARRSAWMRLMFLPFMVSPVCVSAGVLLLPQWTASLPALLAMAYAL
 330 340 350 360 370 380

5 orf139.pep LAYPFVAKDVLSAWDALPPDYGRAAGLGANGFQACRITFPFLPKPALRRGLTLAAATCV
 orf139a LAYPFVAKDVLSDALPPDYGRAAGLGANGFQACRITFPFLPKPALRRGLTLAAATCV
 390 400 410 420 430 440

10 orf139.pep GEFAATFLFSRFEWQTLTLLIYAYLGRAGEDNYARAMVL
 orf139a GEFAATLFXSRXEWQTLTLLIYAYXGRAGXDNYARAMVLTLLAAAFALGXFLLDGGEGG
 450 460 470 480 490 500

The complete length ORF139a nucleotide sequence <SEQ ID 577> is:

1 ATGGATGGAC GGCCTTGGGC GGTATGGSGT GCTTTGGCC TCGTGCCTTC
 51 GGCCTTTTTG GCGGCAATGG TCGTTGCGCG TTTGTCGGCG GTGCGGCGGT
 101 ATGACGGTTT GCGCTGGCGC GCGGTCCTGT CATATCGCTA TATGCTCAAA
 151 CGTTTGGCGT GAGCGGTATG TCGAGCAGCG GCAACCTGCT TCGTGTGCTT
 201 GCGCTTGGCG GTGCTCTFCG CGTGGTGCT GCGCGCGGT GCGTTTCCGG
 251 CGCGGCTTT GTGCTGCGC CTGCTGATGC TGCCCTTTGT GATGCCCGAC
 301 TTTGCTGGCG GCGTGGCGCT GCTGGCTCTG TCGGGGCGG ACGGCTGTGN
 351 CTGCGCGCGC TGGCAGGATA CGCCGTATCT GTTGTGTAC GGCAATGTGT
 401 TTTTNNACCT TCGTGTGTGT GTCAGGCGCG CATATCAGGG GTTGTGCAA
 451 GTGCTGCGG CACGCTCTCA GACGCGCANG ACATGGGCG CGGGGCGCTG
 501 GCGGCGGTTT TGGGACATTG AATGCGCGT TTTGCGCGG TGCGTTGCGG
 551 GCGGCGTGTG CTTGTCTCT CTTGATTTGT TTTGCGGTT CGGGCTGGCA
 601 TTTGCTGCTG TCGGCGAGCG TTATGCCACG TTCGAAGTGG AAATTTACCA
 651 GTTGTGTCAT TCGAATCTCG ATATGCGCGT TGCTTCGGTG CTNGTGTGGC
 701 TGGTGTNGGG GGTAAACNCG GCGGCGAGGT TGCTGTATGC GTGGTTGCGG
 751 AGGCGCGCGG TTTGGGATAA GGCNCTTTCC CCGTGTATGC CCGCGCGCGC
 801 GCGTGTGCTG GGGGAATATG TGCTNCTGCG GTTTCGGCG CGSGTGTNGT
 851 CTGCTGTGCT CCGTGTCTCT TTTGTCGCAA TTGTGTGAA AGCGTGTGCT
 901 GCGGCGGAAT CTGGCGTGTG GTTATGSAA AGTGAACGT GCGAGCGGT
 951 GTGGAATACT TTGCGCTCT CCGCGCGCGG GCTGTATGC GCGGCGGTT
 1001 TGGGTGTGCT GTATGCGCGG GCGGCGCGCG GGTGCGGCT GATGCGCGGG
 1051 CTGATGTTTT TGCCGTTTAT GGTGTCGCGT GTTGTGTGTT CGGCGCGCGT
 1101 GCTGCTGCTT NATCGCGAGT GGACGCGTTC GTTGCCGCTG CTGCTGGGCA
 1151 TGATGCGCGT GCTGGCGTAT CGGTTTGTG CAAAGATGT TTTATCAGCC
 1201 TGNGATGCAC TGCCGCGCGA TTACGCGAGG GCGGCGCGG GTTGGGTGCT
 1251 AAACGCGCTT CAGACGGCAT CGCGCATCAC GTTCCGCTC TTGAACCGCG
 1301 CGTTGCGCGG CGGTCTGACT TTGCGGCGCG CAACCTGCGT GGGCGAATTT
 1351 GCGGCAACCT GTTONTGTC CGGTCTCGAG TGGCAGACGC TGACAGCTTT
 1401 GATTATGCT TATNTGGGAC GCGCGGTTGA NGATATATC GCGCGGGGCA
 1451 TGGTGTGCTG ATTGCTGTG GCGGCGTTCT CGCTGGGTAT NTTCTGCTG
 1501 TTGACGCGG GCGAAGCGG AAAACGAGC GAAACGTAT AA

This encodes a protein having amino acid sequence <SEQ ID 578>:

1 MDGRRWAVWG AFALLPSAFL AAMVVAFLWA VAAYDGLAWR AVLSDAYMLK
 51 RIANTVFQAA ATCVLVPLG VFVWVWLRL AEPGRALVLR LLMLEFVMT
 101 LVAGVVLVLA RGADGLXWPG WQDFPFLLY GNVEKFLVR VRAAYQFVQ
 151 VPAARLCTAX TLGAGWRRF WDIEMVPLRP WLAGGVCLVF LYCESGFLGL
 201 LLLGGSRYAT VEVEIYQLWM FELDMVASV LVWLVGXVTA AAGLLYAWFG
 251 RRAVSDKAVS PVPMPSPQVS GEYVLIAFA AVXSVCCFLX LLAIVVKVMS
 301 AGESWRVIME SETWQAVVNT XRFSAAYVA AAVLGVVYAA AARRSAMRG
 351 LMLFPMVSP VCVSAGVLLL XPOWTASLEL LLAMYALAY PFVAKDVLSA
 401 XDALPPDYGR AAGLGANGF QACRITFEL LKPALRRGLT LAAATCVGEF
 451 AATLFXSRXE WQTLTLLIYA YXGRAGXDNY ARAMVLTLLL AAFALGXFL
 501 LDGEGGKRT ETL*

ORF139a and ORF139-1 show 96.5% homology over a 514aa overlap:

orf139a.pep MDGRRWAVWGAFALLPSAFLAAMVVAFLWAAVAYDGLAWRAVLSDAYMLKRLAWTVFQAA
 orf139-1 MDGRRWVWGAFALLPSAFLAVMVAFLWAAVAYDGLAWRAVLSDAYMLKRLAWTVFQAA

	orf139a.pep	ATCVLVLPFGVPAWVLARLAFPGRALVRLRLMLPFVMPMTLVAGVGVLALEFGADGLXWRG
	orf139-1	ATCVLVLPFGVPAWVLARLAFPGRALVRLRLMLPFVMPMTLVAGVGVLALEFGADGLLWRG
5	orf139a.pep	WQDTPYLLLYGNVFFKLPVLVRAAYQGFVQVPAARLQTAXTLGAGAWRRFWDIEMPVLRP
	orf139-1	RQDTPYLLLYGNVFFNLPLVLRAAYQGFVQVPAARLQTARTLGAGAWRRFWDIEMPVLRP
10	orf139a.pep	WLAGGVCLVFLYCFSGFGLALLGGSRVATVEVEIYQLVMFELDMAVASVLVWLVLXGVT
	orf139-1	WLAGGVCLVFLYCFSGFGLALLGGSRVATVEVEIYQLVMFELDMAVASVLVWLVLGVT
15	orf139a.pep	AAGLLYAWFGRRAVSDKAVSPVMPSPQSVGEYVLLAFAAAVXSVCCFLXLLATVVKAWS
	orf139-1	AAGLLYAWFGRRAVSDKAVSPVMPSPQSVGEYVLLAFAAAVXSVCCFLXLLATVVKAWS
20	orf139a.pep	AGESWRVLMSESTWQAVWNTXRFSAAAVYAAAVLGVVVYAAAARRSAMWRGLMFLPFMVSP
	orf139-1	AGESWRVLMSESTWQAVWNTLRFSAAAVYAAAVLGVVVYAAAARRSAMWRGLMFLPFMVSP
25	orf139a.pep	VCVSAGVLLXLPQWTASLPDLLAMYLALAYPFVAKDVLXAKDALPPDYGRAAGLGANGF
	orf139-1	VCVSAGVLLXLPQWTASLPDLLAMYLALAYPFVAKDVLXAKDALPPDYGRAAGLGANGF
30	orf139a.pep	QTACRITFPLKLPALRRGLTAAATCGVEFAATLFLSRXEWQTLTTLIYAYXGRAGXNDY
	orf139-1	QTACRITFPLKLPALRRGLTAAATCGVEFAATLFLSRPEWQTLTTLIYAYLGRAGEDNY
35	orf139a.pep	ARAMVLTLLAAAFALGXFLLLDGGEGGKRTETLX
	orf139-1	ARAMVLTLLAAAFALGIFLLDGGEGGKGTETLX

Homology with a predicted ORF from *N.gonorrhoeae*

ORF139 shows 95.2% identity over a 189aa overlap with a predicted ORF (ORF139ng) from

N.gonorrhoeae:

35	orf139.pep	AWGAGESWRVLMSESTWQAVWNTLRFSA	30
	orf139ng	QSVGEYVLLAFSAVLSVCCFLPLSAIVKVASAGESRRVLMSESTWQAVWNTLRFSA	327
40	orf139.pep	VYAAAVLGVVYAAPARRSAMWRGLMEXPFMVSPVCVSAGVLLLYPQWTASLPDLLAMYL	90
	orf139ng	VFAAAVLGSVVYAAAARRLVMMRGLVFLPFMVSPVCVSAGVLLLYPGWTASLPDLLAMYL	387
45	orf139.pep	LAYPFVAKDVLXAKDALPPDYGRAAGLGANGFQACRITFPLKLPALRRGLTAAATCV	150
	orf139ng	LAYPFVAKDVLXAKDALPPDYGRAAGLGANGFQACRITFPLKLPALRRGLTAAATCV	447
50	orf139.pep	GEFAATLFLSRPEWQTLTTLIYAYLGRAGEDNYARAMVL	189
	orf139ng	GEFAATLFLSRPEWQTLTTLIYAYLGRAGEDNYARAMVLTLLSAFAVCIFLLDNGEGG	507

The complete length ORF139ng nucleotide sequence <SEQ ID 579> is predicted to encode a protein having amino acid sequence <SEQ ID 580>:

55	1	MDGRCAWVRG	AFSLLPSAFL	AVMVVAPLWA	VAAYDGLAWR	AVLSDAYMLK
	51	RLANTVFQAA	ATCVLVLQ	VFWAVVLARL	AFPGRALVLR	LLMLPFVMPMT
	101	LVAGVGVLALE	FGADGLLWRG	RQDTPYLLLY	GNVFNLFVL	VRAAYQGFQ
	151	VPAARLQART	TLGAGAWRRF	WDIEMPVLR	WLAGGVCLVF	LYCFSGFGLA
	201	LLGGSRVAT	VEVEIYQLVM	FELDMAGASA	LVLVLGVT	AAGLYRAWG
	251	RRAVSDKAVS	PVMPSPQSV	GEYVLLAFSV	AVLSVCCFLP	LSAIVVKAWS
	301	AGESRRVLM	SETWQAVWNT	LRFSAAAVFA	AAVLGVVYAA	AARRLVMMRG
60	351	LVFLPFMVSP	VCVSAGVLL	YPGWTSASLP	LLAMYLALAY	PFVAKDVLXA
	401	WDALPPDYGR	AAAGLGANGF	QTACRITFPL	LKLPALRRGLT	LAATCVGEF
	451	AATLFLSRPE	WQTLTTLIYA	YLGRAGEDNY	ARAMVLTLL	SAFAVCIFLL
	501	LDNGEGGKRT	ETL*			

Further work revealed a variant gonococcal DNA sequence <SEQ ID 581>:

```

1  ATGGATGGAC  GGTGTGGGCG  GGTACGGGGT  GCTTTTCC  TGCTGCCTTC
51  GCGCTTTTTT  GCGGTAATGG  TCGTTGCGCC  TTTGTGGGG  GTGGCGGGT
101  ATGACGGTTT  GCGGTGCGCG  GCGGTGCTGT  CGGATGCCTA  TATGCTCAAA
151  CGTTTGGCGT  GGAACGGTGT  TCAGGCGGCG  GCAACCTGTG  TGCTGTGCT
201  CGCTTTGGCG  GTGCTGTGCT  CGTGGGTGCT  GCGCGGCGTG  GCGTGTGCT
251  GGGGGGCTTT  GGTGCTGCGG  CTGCTGATGC  TGCCTGTGCT  GATGCCACAG
301  CTGGTGGCGG  GCGTGGCGGT  GCTGCTGTG  TTGCGGGGG  ACGGGCTGTT
351  GTGGCGGCGG  GCGTGGCGGT  GCGCTATCT  GTTCTGTAC  GCGAATGTGT
401  TTTTTCGCGG  GCGCTGTGTT  GTACGGGCG  CGTATCAGGG  GTTTGCTCAA
451  GTGCTTCGCG  CACGGCTTCA  GACGGCACGG  ACGTGTGGCG  CGGGGCGGTG
501  CGGCGCGTGT  TGGGACATTG  AATGCGCGT  TTTGCGCCG  TGGCTTGGCG
551  CGGCGGTGTG  CTTTGTCTTC  CTGTATTGTT  TTTGCGGGT  CGGGTGTGCA
601  TTGCTGTGTG  GCGGCAGCGG  TTATGCCACG  GTCGAAGTGG  AAATTTACCA
651  GTTGTGTATG  TTCGAATCTG  ATATGCGGGG  GGCTTCGGCG  CTGCTGTGGC
701  TGGTGTGTGG  GGTAAACGGG  GCGGCAGGGT  TTTGCGCCG  GTGCTTCGCG
751  AGGCGCGCGG  TTTGCGATAA  GCGCGTTTCC  CCGGTGATGC  CTGCGCGCCG
801  GCAATCGGTG  GGGGAATATG  TATTGCTGGC  ATTTTCGGTG  GCGGTGTGTG
851  CCGTGTGCTG  CTTGTTTCC  TTTGCGGCAA  TTGTGTGAA  AGCGTGGTGC
901  GCGGCGGAAT  GCGGCGGTG  GTTAATGAA  AGTGAAACGT  GCGACGAGT
951  GTGGAATACT  ttGCGCTTTT  GCGCGGCGCG  GGTGTTTGGC  GCGGCGGGT
1001  TGGGTGTGGT  GTATGCGGCG  GCGCGCGCGG  GCGTGTGTG  GATGCGGGTA
1051  CTGGTGTGTT  TACCGTTTAT  GGTGCTGGCG  GTGCTGTTT  CGGCGGCTT
1101  CGACGGCTTC  TATCCGCGGT  GTTACGGCTG  GTTACGGCTG  CPGCTGGCGA
1151  TGTATGCGCT  GCTGCGGCTT  CCGTGTGTG  CAAGAATGT  TTTATCGGCC
1201  TGGGATGCAC  TCGCGCGGGA  TTACGCGCAG  GCGGCGGCG  GTTTGGGCGC
1251  AAACGCGCTT  CAGACGCGAT  GCGTATCAC  GTTCCCCCT  TTGAACCGCG
1301  CGTTGCGCGG  CGGTCTGACT  TTGCGGCGCG  CGACGTGTG  GCGCGAATTT
1351  GCGGCAACCT  TGTCTCGTGC  GCGTCCGGA  TGCGACAGT  TGACGACTTT
1401  GATTTATGCC  TATTGTTGGG  GTGCGGGTGA  GGACAATTAT  GCGCGGGCAA
1451  TGGTGTGTAC  ATTGCTGTGT  TCGGCAATTG  CGGTGTGCAT  TTTCTGCTG
1501  TTGGACAACG  GCGAAGGCGG  aaaACGGACG  GAACGTTAT  AA

```

This corresponds to the amino acid sequence <SEQ ID 582; ORF139ng-1>:

```

1  MDGRCAVVRG  AFSLLPSAFL  AVMVVAFLWA  VAAYDGLAWR  AVLSDAYMLK
51  RLAWTVFQAA  ATCVLVPLPG  VPVAVVLARL  AFPGRALVLR  LIMLPFVMT
101  LVAGVGVLLA  FGADGLLWRG  RQDTPYLLLY  GNVPFNLVPL  VRAAYQGFGA
151  VPAARLQTA  TLGAGAWRRF  WDIEMPVLRP  WLAGGVCLVF  LYCFSGFLA
201  LLLGGSRYAT  VEVEIYQLVM  FELDMAGASA  LWLVLVGVT  AAGLLYAWFG
251  RRAVSDKAVS  PVMPSPPQSV  GEYVLLAFSV  AVLSVCLFFP  LSAIVVKAMS
301  AGESRRVLME  SETWQAVWNT  LRFSAARVFA  AARLVWVRA  AARLVWVRA
351  LVFLPFMVSE  VCVAGVLLY  YFGWATLDEL  PFGADVLSA  AADLPEFGR
401  WDLPEFGR  WADLAGANGF  QACRITTFPL  LKPARRLGT  LAATCVGGE
451  NATLFSRPE  WQTITTLIYA  YLGRAGEDNY  ARAMVLTLLL  SAFAVCIFLL
501  LDNGEGGKRT  ETL*

```

ORF139ng-1 and ORF139-1 show 95.9% identity over 513aa overlap:

```

orfl39ng      MDGRCAVVRGAFSLLPSAFLAVMVVAFLWAVAAYDGLAWRAVLSDAYMLKRLAWTVFQAA
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
orfl39-1      MDGRVWVWGAFALLPSAFLAVMVVAFLWAVAAYDGLAWRAVLSDAYMLKRLAWTVFQAA
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |

50 orfl39ng      ATCVLVPLGVPVAVVAVLRLAFLPGRALVLRLLMLPFVMTLVAGVGVLLAFLGADGLLWRG
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
orfl39-1      ATCVLVPLGVPVAVVAVLRLAFLPGRALVLRLLMLPFVMTLVAGVGVLLAFLGADGLLWRG
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |

55 orfl39ng      RQDTPYLLLYGNVPFNLVPLVRAAYQGFGAQPAAARLQARTLGAAGWRRFWDIEMPVLRP
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
orfl39-1      RQDTPYLLLYGNVPFNLVPLVRAAYQGFGVQVPAARLQARTLGAAGWRRFWDIEMPVLRP
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |

orfl39ng      WLAGGVCLVFLYCFSGFGLALLLGGGSRYATVEVEIYQLVMFELDMAGASAVLWLVGVT
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
60 orfl39-1      WLAGGVCLVFLYCFSGFGLALLLGGGSRYATVEVEIYQLVMFELDMASAVLWLVGVT
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |

orfl39ng      AAGLLYAWFGRRRAVSDKAVS PVMPSPPQSVGEYVLLAFSAVLSVCLFFPLSAIVVKAMS
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
orfl39-1      AAGLLYAWFGRRRAVSDKAVS PVMPSPPQSVGEYVLLAFSAVLSVCLFFPLSAIVVKAMS
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |

```


orf139ng AGESRRVILMESETWQAVWNTLRFSAAVFAAAVLGVVYAAAAARRLVWVRGLVFLPFMVSP
 orf139 AGESRRVILMESETWQAVWNTLRFSAAVFAAAVLGVVYAAAAARRLVWVRGLVFLPFMVSP
 orf139ng VCVSAGVLLLYPGWTASLPLLLAMYALLAYPFVAKDVL~~SAWDALPPDYGRAAAGL~~GANGF
 orf139-1 VCVSAGVLLLYPQWTASLPLLLAMYALLAYPFVAKDVL~~SAWDALPPDYGRAAAGL~~GANGF
 orf139ng QTACRITFLLKPALRRGLTAAATCVGEFAATLFLSRPEWQTLLTLLIYAYLGRAGEDNY
 orf139-1 QTACRITFLLKPALRRGLTAAATCVGEFAATLFLSRPEWQTLLTLLIYAYLGRAGEDNY
 orf139ng ARAMVLTLLLSAFAVCI~~FLLLDNGEGGKRT~~ETFL
 orf139-1 ARAMVLTLLLSAFALGIF~~LLLDGEGGKQT~~ETFL

Based on the presence of a predicted binding-protein-dependent transport systems inner membrane component signature (underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 70

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 583>:

1 ATGGACGGCT GGACACAGAC GCTGTCGGG CAAACCTGT TGGGCATTTC
 51 GCGCGCGGCA ATCATCCTCA TTCTGATTTT AATCGTCAAA TTCGCGATCC
 101 ACAGCGCTGCT GACACTGGTC ATCGTCAGCC TGCTGACGGC TTGGCAACCC
 151 GGTTCGCCCA CAGGACGATC TGTCACAGAC ATACTGTGCA AAAACTTCGG
 201 CGGACGCGTC GCGCGCTGG CGCTCTGCT GCGCCTGGGC GCGATGCTCG
 251 AACGTTTGGT C...

This corresponds to the amino acid sequence <SEQ ID 584; ORF140>:

1 MDGWTQTLSA QTLIGISAAA IILILILIVR FRIHALTLV IVSLLTALAT
 51 GLPTGSIVKD ILVKNFGTLL GVVALLVGLG AMLERIV..

Further work revealed the complete nucleotide sequence <SEQ ID 585>:

1 ATGGACGGCT GGACACAGAC GCTGTCGGG CAAACCTGT TGGGCATTTC
 51 GCGCGCGGCA ATCATCCTCA TTCTGATTTT AATCGTCAAA TTCGCGATCC
 101 ACAGCGCTGCT GACACTGGTC ATCGTCAGCC TGCTGACGGC TTGGCAACCC
 151 GGTTCGCCCA CAGGACGATC TGTCACAGAC ATACTGTGCA AAAACTTCGG
 201 CGGACGCGTC GCGCGCTGG CGCTCTGCT GCGCCTGGGC GCGATGCTCG
 251 GACGTTTGGT CGAAACATCC GCGCGCGATC AGTCGCTGGC GGACGCGCTCG
 301 ATCGGATGTC TCGCGCAAAA ACAGCGACCG TCAGCGCTGG GCGTTGCTCG
 351 GCTGATTTTC GGCTTCCCGA TTTTCTTCCA TCCCGGACTA ATCGTCATCG
 401 TGCCCATCGT GTTCGCCAACC GCACGGGCGA TGAACACGAA ATCGTCGCCC
 451 TTCGCGCTTG CCTCCATCGG CGCATTTTCC GTCAGGCAAG TCTTCTCGGC
 501 GCGCCATCCG GCGCCGATTG CCGCTTCCGA ATTTTACGCG CGCAACATCG
 551 GCGCAAGTTT GATTTTGGGT CTGCGGACCG CTTTCAATCA ATGATATTTC
 601 AGCGGCTATA TGCTCGGCAG AGTGTTCGGG CGCACCATCC ATGTTCCCGT
 651 TCCCGAACTG CTCAGCGGGC GCACGCAAGA CAACGACCTG CGCAAGAGAC
 701 CTGCGAAAGC AGGAACGGTC GTCCGCATCA TGCTGATTCC CATGCTGCTG
 751 ATTTTCTCTGA ATACGCGCAT ATCGGCCCTC ATCGAGGAAA AACTCGTAAG
 801 TCGCGACGAA ACCTGGGTTT AGACGCGCAA AATAATCGGT TCGACACCGA
 851 TCGCCCTTCT GATTTCCGTA TTTGTCGCAC TGTTTGTCTT GGGACGCAAA
 901 CGCGCGGAAA GCGGACGCG GTTGAAAAAA ACGGTGGACG CGGCACCTCG
 951 CCGCGCTCTG TCGGTGATTG TGATTACCGG CGCGGCGCGT ATGTTTCGGG
 1001 GCGTTTTCGG CGCTTCCGGC ATCGGCAAGG CACTGCGCGA CAGCATGGGG
 1051 GATTTCGGCA TTCCCGTCTT TTTGGCTCTG TTCTTCTGCG CTTTGGCATC
 1101 GCGTATCGCG CAAGGTTGGG CAACCTCGCG CTGACACACC GCGCGCGCGC
 1151 TGATGCTCTC TGCGGTTGCC GCGCGCGGCT TTACCGACTG GCAGCTCGCC

1201 TGTATCGTAT TGGCAACGGC GGCAGGTTCC GTCGGTGCA GCCACTTCAA
 1251 CGACTCCGGC TTCTGGCTGG TCGGCCGCTC CTGGACATG GACGTACCGA
 1301 CCACGCTGAA AACCTGGACG GTCAACCAA CCCTCATCGC ACTCATCGGC
 1351 TTTGCCTGT CCGCACTGCT GTTCGCCATC GTCTGA

5 This corresponds to the amino acid sequence <SEQ ID 586; ORF140-1>:

1 MDGWTQTLQA QTLIGISAAA IILILILIVK FRIHALLTLV IVSLLTALAT
 51 GLPTGSIVND ILVKNFGGTL GGVALLVGLG AMLGRVLVTS GGAQSLADAL
 101 TRMFGKEKRAP FALGVASLIF GFPIFFDAGL IVMLPFI VAT ARMKQDVLP
 151 FALASIGAFS VMHVFLPPH GFIAASEFYG ANIQGVLLIG LPTAFITWYF
 201 SGYMLGKVLG RTIHVPPEL LSGGTQDNLD KPEPAKAGTV VAIMLIPMLL
 251 IFLNTGVSAL ISEKLVSADL TWVQTAKIIG STPIALLISV LVALFVLGRK
 301 RGEESGALEK TVDGLAPVPC SVILITGAGG MEGGVLRASG IGKALADMSA
 351 DLGIPVLLGC FLVALALRIA QGSATVALTT AAALMAPAVA AAGFTDWQLA
 401 CIVLATAAGS VGCSEFNDG FVLVGRLLDM DVPTTLTKWT VNQTLIALIG
 451 FALSALLFAI V*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF140 shows 95.4% identity over a 87aa overlap with an ORF (ORF140a) from strain A of *N.*

meningitidis:

		10	20	30	40	50	60
20	orf140.pep	MDGWTQTLQAQTLIGISAAA	IILILILIVK	FRIHALLTLV	IVSLLTALAT	GLPTGSIVKD	
	orf140a	MDGWTQTLQAQTLIGISAAA	IILILILIVK	FRIHALLTLV	IVSLLTALAT	GLPTGSIVND	
		10	20	30	40	50	60
25			70	80			
	orf140.pep	ILVKNFGGTLGGVALLVGLG	AMLERLV				
	orf140a	ILVKNFGGTLGGVALLVGLG	AMLERLV	ETSGGAQSLADAL	TRMFGKEKRAP	FALGVASLIF	
		70	80	90	100	110	120
30							

The complete length ORF140a nucleotide sequence <SEQ ID 587> is:

1 ATGGACGGCT GGACACAGAC GCTGTCCCGC CAAACCCGTG TGGGCATTTC
 35 51 GCGCGCGGCA ATCATCTCTA TTCTGATTTT AATCGTCAA TTCCGATCC
 101 ACGCGCTGCT GACACTGTGC ATCGTCAGCC TGCTGACGCC TTGGCAACC
 151 GGTTTGCCCA CAGGCAGCAT TGTCACAGAC CACTCGGTCA AAACTCTCGG
 201 CGGCACGCTC GCGCGCGTGG CGCTTCTGGT CGGCCTGGGC GCGATGCTCG
 251 GACGTTTGGT CGAAACATCC GCGCGCGCAC AGTGCTGGC GGCACGCGTG
 301 ATCCGGATGT TCGCGCAAAA ACGCGCACOG TTCGCGTGG CGGTTGCGCT
 351 GCTGATTTTC GCGCTCCCGA TTTTCTCGA TCGCGACTA ATCGTCATGC
 401 TGCCCATCGT GTTCGCCACC GCACGGCGCA TGAAACAGGA CGTACTGCC
 451 TTGCGCGTTG CTTCCATCGG CGCATTTTCC GTCATGACG TCTTCTGCC
 501 GCCCATCCG GCGCGGATTG CCGCTTCCGA ATTTTACGCG CGCAACATCG
 551 GCCAAGTTTT GATTTTGGGT CTGCGCACOG CTTTCATCAC ATGGTATTT
 601 AGCGGCTATA TGCTCGSCAA AGTGTGGGG CGCAACATCC ATGTTCCCGT
 651 TCCCGARCTG CTCAGCGCGG GCACGCAAG CACGACCTG CGCAAGAGAC
 701 CTCGCAACG AGGAACGTC GTGCGACTCA TCGTGATCC CATGCTGCT
 751 ATTTTCTGTA ATACCGCGGT ATCGGCGCTC ATCAGCGAAA AACTCTAAG
 801 TCGGACGAAA ACCTGGGTTT AGACGCGAAA AATAATCGT TCGACACGA
 851 TCGCCCTTCT GATTTCGTA TTGGTCCGAC GTTTTGTCTT GGGACGCGAA
 901 CGCGCGGAAA CGCGCAGCGC GTTGGAAAAA ACCGTGGACG GCGCATCTGC
 951 CCCGCTCTGT TCCGTGATTC TGATTACCGC CGCGGGCGGT ATGTTTGGCG
 1001 GCGTTTTCGG CGCTTCCGCG ATOGGCAAG CACTCGCGA CAGCATGGCG
 1051 GATTTGGGCA TTCCGTCCTT TTTGGGCTGT TTTCTGTGCT CTTTGCACCT
 1101 GCGTATCGCG CAAGGTTTCG CAACCGTCGC CGCCACACC CGCGCGCGCG
 1151 TGATGGCTCC TGCCGTTGCC GCGCGCGGCT TTACCGACTG GCAGCTCGCC
 1201 TGATTCGTAT TGGCAACGGC GCGAGGTTGC GTTCGGTTGA GCGACTTCAA
 1251 CGACTCCGGC TTCTGGCTGG TCGGCGCGCT CTGGACATG GACGTACCGA
 1301 CCACGCTGAA AACCTGGAOG GTCAACCAA CCCTCATCGC ACTCATCGGC
 1351 TTTGCTGTG CCGCACTGCT GTTCGCCATC GTCTGA

This encodes a protein having amino acid sequence <SEQ ID 588>:

```

1 MDGWQTQLSA QTLGISAAA IILILILIVK FRIHALLTLV IVSLLTALAT
51 GLPTGSIVND VLVKNFGGTL GGVALLVGLG AMLGRIVETS GGAQSLADAL
101 IRMFGEKRAP FALGVASLIF GFPIFFDAGL IVMLPIVFAT ARRMKQDVLV
151 FALASIGAFS VMHVFLPPEH GPTAASEFYG ANIGQVLIIG LPTAFITWYF
201 SGYMLGKVLG RTHVVPPEL LSGGTQDNDL KKEPAKAGTV VAIMLI PMLL
251 IFLNTGVSAL ISEKLVSADE TWVQTAKIIG STPIALLISV LVVALFVLGRK
301 RGESGSALEK TVDGLAPVC SVILITGAGG MFGGVLRASG ICKALADMSA
351 DLGIPVLLGC FLVALALRIA QGSATVALTT AAALMAPAVA AAGTTDWQLA
10 CIVLATAAGS VGCSEFNDG FWLVGRLLDM DVPTTLKTWT VNQTLIALIG
451 FALSALLFAI V*

```

ORF140a and ORF140-1 show 99.8% identity over a 461aa overlap:

```

15 orf140-1.pep MDGWQTQLSAQTLGISAAAIIILILILIVKFRIHALLTLVIVSLLTALATGLPTGSIVND 60
orf140a MDGWQTQLSAQTLGISAAAIIILILILIVKFRIHALLTLVIVSLLTALATGLPTGSIVND 60

20 orf140-1.pep ILVKNFGGTLGGVALLVGLGAMLGRIVETS SGGAQSLADALIRMFGEKRAPFALGVASLIF 120
orf140a VLVKNFGGTLGGVALLVGLGAMLGRIVETS SGGAQSLADALIRMFGEKRAPFALGVASLIF 120

25 orf140-1.pep GFPIFFDAGLIVMLPIVFATARRMKQDVLFPFALASIGAFSVMHVFLPPEHGPPIAASEFYG 180
orf140a GFPIFFDAGLIVMLPIVFATARRMKQDVLFPFALASIGAFSVMHVFLPPEHGPPIAASEFYG 180

30 orf140-1.pep ANIGQVLIIGLPTAFITWYFSGYMLGKVLGRTHVVPPELLSGGTQDNDLPKEPAKAGTV 240
orf140a ANIGQVLIIGLPTAFITWYFSGYMLGKVLGRTHVVPPELLSGGTQDNDLPKEPAKAGTV 240

35 orf140-1.pep VAIMLI PMLLIFLNTGVSALISEKLVSADETWVQTAKIIGSTPIALLISVLVALFVLGRK 300
orf140a VAIMLI PMLLIFLNTGVSALISEKLVSADETWVQTAKIIGSTPIALLISVLVALFVLGRK 300

40 orf140-1.pep RGESGSALEKTV DGLAPVCSVILITGAGGMFGVLRASGIGKALADMSADLGPVLLGC 360
orf140a RGESGSALEKTV DGLAPVCSVILITGAGGMFGVLRASGIGKALADMSADLGPVLLGC 360

45 orf140-1.pep FLVALALRIAQGSATVALTTAAALMAPAVAAAGFTDWQLACIVLATAAGSVGCSEFNDG 420
orf140a FLVALALRIAQGSATVALTTAAALMAPAVAAAGFTDWQLACIVLATAAGSVGCSEFNDG 420

50 orf140-1.pep FWLVGRLLDMVDVPTTLKTWTVNQTLIALIGFALSALLFAIV 461
orf140a FWLVGRLLDMVDVPTTLKTWTVNQTLIALIGFALSALLFAIV 461

```

45 Homology with a predicted ORF from *N.gonorrhoeae*

ORF140 shows 92% identity over a 87aa overlap with a predicted ORF (ORF140ng) from *N.gonorrhoeae*:

```

50 orf140.pep MDGWQTQLSAQTLGISAAAIIILILIVKFRIHALLTLVIVSLLTALATGLPTGSIVND 60
orf140ng MDGWQTQLSAQTLGISAAAIIILILIVKFRIHALLTLVIVSLLTALATGLPTGSIVND 60

orf140.pep ILVKNFGGTLGGVALLVGLGAMLERLV 87
orf140ng VLVKNFGGTLGGVALLVGLGAMLGRIVETS SGGAQSLADALIRMFGEKRAPFAPGVASLIF 120

```

55 The complete length ORF140ng nucleotide sequence <SEQ ID 589> was predicted to encode a protein having amino acid sequence <SEQ ID 590>:

```

1 MDGRQTQLSA QTLGISAAA IILILILIVK FRIHALLTLV IASLLTALAT
51 GLPTGSIVND VLVKNFGGTL GGVALLVGLG AMLGRIVETS GGAQSLADAL
101 IRMFGEKRAP FAPGVASLIF GFPIFFDAGL IVMLPIVFAT ARRMKQDVLV

```

5

151	FALASVGAFS	VMHVFLPHF	GPIAASEFYG	ANIGQVLLIG	LPTAFITWML
201	SGYMLKRVLG	RATHVPVFEL	LSGGTQSDSD	PKPEKAGTIV	VAVMILPML
251	IFLNTGVSYAL	ISEKLVSADL	TWQVTKQMIG	SPVALLISLV	LAALLVLGRK
301	RGGSSGTLK	TVFGALAPAC	VLITLTGAGG	MFGGVLRASS	IGKALADASA
351	DLGILPVLLG	FLVALALRIA	QGSATYATGT	AAALMAPAVA	AAGFTWQDLA
401	CIVLATAAGS	VGCSEHFNDS	FWLVGRLSDM	DVPTTLKTWT	VNQTLLAFIG
451	FALLSLFAFI	V*			

Further work revealed a variant gonococcal DNA sequence <SEQ ID 591>:

1	ATGGACGAGCC	GACACAGACG	GCTGTCCGGG	CAATCGTTGT	TGGCGATTTT
51	GGCGGCGGCA	GTACATCTCA	TCCTGATTTT	AAGAGCTCAA	TTCCGATCAT
101	GGCGCGTCTT	GACATCTGTG	ATCGCGACCC	TGCTACGCGG	TTTGCGAACCC
151	GCTTTGGCCA	CAGCGAGCAT	GTCTCAACAG	TGACTGTGCA	AAAATCTTCG
201	CGGCACGCTC	GGCGCGCTGG	CGCTTCTGCT	CGGCTTGGGC	GCAATGCTCG
251	GACGTTTGGT	AGGAACAATC	GGCGGCGGCA	AGTCTGTGGC	GGACGCGTGT
301	ATCCGGATGT	TGGCGGAAAA	ACGGCGACCG	TTTCGCTCGG	CGGTTGCCCTC
351	GCTGATTTTC	GCTGCTCCGA	TTTTCTTCGA	TGCGCGACGA	ATCGCTCATC
401	TGCCCATCGT	ATCTGCCGCC	CGACGGCGCA	TGCAAGAACGA	CGTATCGCCCT
451	TTGCGCGCTG	CTCTCGCTGG	CGGATTTTCC	GTCATGCAGC	TCCTTCGTGG
501	GCCCCATTCG	GGCCCGGATG	CGGCTTCCGA	ATTTTACGCG	CGGAACATCTG
551	CGCAGSTTCC	GATTTTGGGT	GTCCGCGACG	CTTCTCATAC	ATGGTATCTC
601	AGCGCGTATA	TGCTTGGGCA	AGCTGTGGGG	CGTCCGATCC	ATCTGCGCTG
651	TCGCGGCGCG	CGACGCGCGC	GACGACGCGC	CGACGACCCG	CGCGAAGAAC
701	TCGCAAGAA	AGCGACGCTC	GTGCCCTCTA	TGCTGATTCG	CATGCTGCTG
751	ATTTTCTCTGA	ATACCGGCGCT	ATCCAGCCCTC	ATCAGCGAAA	AACTCGTFAG
801	ACGCGACGAA	AGCTTGGGTT	AGACGCGCAA	AATGATCGGT	TCGACACCTG
851	TGCGCCCTCT	GATTTCCGTA	TGTGGCGCAC	TGTGTGTTCT	GGGACGCAAA
901	CGCGGCGAAA	CGGGGACGAC	TCGAGAAAAA	ACCTGTGGAG	CGCAGCTCGT
951	CGCCGCGCTG	CTCGGATGAT	TGATTACGCG	CGCGGCGCGT	ATGTTGCGGG
1001	CGTTTTTGGG	CGCTTCCGCG	ATCTGCAAGG	CAGCTGCGGA	CAGCATGGGG
1051	GATTTGGGCA	TTCCGCTGCT	TTTGGGCTGC	TTCTCTGTGC	CTTGGGACAT
1101	CGGTATCGCG	GAGGCTTCGG	CACCGCTCGC	CTCTGACCAC	CGCGCGCGCG
1151	TGATGGCTCC	TGCGCTTGCC	CGCGCGCGCT	TTACCGACTG	CGAGCTGGCC
1201	TGTATCGTAT	TGGCAACGCG	CGTAGGTCGT	CGGCTGCTGA	CGCAGTCTGC
1251	CGACGCTGTA	CTCTGCTGCT	TGGCGCGGCA	CTTATGATCA	GAATCTGCTG
1301	CCACGCTGTA	CAACCTGACG	GTCACCAAAA	CCCTCATGCG	ATTTCATCGG
1351	TTTGGCTTGT	AGCGCTGCT	TTTGGACATC	GCTGCA.	

This corresponds to the amino acid sequence <SEO ID 592: ORF140ng-1>:

```

1 MDGRTQTLSA QTLIGISAAA ILLILILVVK FRIRALLTV IASLLTALAT
51 GLPTFGTIVND VLIVNFGFTGL GQVALISLVG AMRLAGLTVS GGAQSLADAL
40 101 IRMFCEKRAP FAFVGVASLIF GPFIIFDAGL IVMLPIVFAT ARRMKQOVLE
151 FALASVGAFS VMHVFLPHP GPEIAAEFFG ANIGOVILIV LPTAFITWYF
201 SCYMLGKVLG RAIHVVPVEL LSGGTQDSDP PKEPAKAGTV VAVMLIPMLL
251 IFLMTGVSAI ISEKLYSADE TW20TKMID SPVPALLISV LAALVLGKR
45 301 RGESGSTEK TVDALAPAC SVSITGAGG MGGGVLRASG ICKALADMS
351 DLGIPVLVLC FLVAGALRIA QQSATVAITF AALMAFAVA AAGFTDWQLA
401 CIVLATAAGS VGCSHFNDSG FNLVGRLLDM DVPTTLKTTV VNQTILIAFIG
451 FALSALLFAI V*

```

ORF140ng-1 and ORF140-1 show 96.3% identity over 461aa overlap:

```

50 orf140ng-1.pep MDGRTQTLSAQTLIGISAAAIIILLILVVKFRIRALLTVLIASLLTALATGLPTGSSVND
orf140-1 MDGWTQTLSAQTLIGISAAAIIILLILVVKFRIRALLTVLSLLTALATGLPTGSSVND
orf140ng-1.pep VLVKNFGTIGGVALLVVLGAMGLRIVETSGGAQSLADALIRMFCEKRAPFAGVASLIF
55 orf140-1 ILVKNFGTIGGVALLVVLGAMGLRIVETSGGAQSLADALIRMFCEKRAPFAGVASLIF
orf140ng-1.pep GPFIIFDAGLIVMLPIVFATARRMKQOVLPFALASVGAFSVMHVFLPHPGPIAAAEFFYG
orf140-1 GPFIIFDAGLIVMLPIVFATARRMKQOVLPFALASVGAFSVMHVFLPHPGPIAAAEFFYG
60 orf140ng-1.pep ANIGOVILILGVLPTAFITWYFSGYMLGKVLGRAIHVVPPELLSGGTQDSDPKEPAKAGTV
orf140-1 ANIGOVILILGVLPTAFITWYFSGYMLGKVLGRTIHVVPPELLSGGTQDSDPKEPAKAGTV

```

```

5  orf140ng-1.pep  VAVMLIPMLLI1FLNTGVSALISEKLVSADETWVQ2TAKMIGSTFVALLISVLAALLVLGRK
   orf140-1        VAIMLIPMLLI1FLNTGVSALISEKLVSADETWVQ2TAKIIGSTFIALLISV3LV4ALFVLGRK
10  orf140ng-1.pep  RGE5SGSTLEKTVDGALAPACSV6ILITGAGGMFGGVL7RASGIGKALADSMADLGI8PVLLGC
   orf140-1        RGE5SGSALEKTVDGALAPVCSV6ILITGAGGMFGGVL7RASGIGKALADSMADLGI8PVLLGC
15  orf140ng-1.pep  FLVALALRIAQGSATV9LTAAALMAPAVA10AAAGFTDWQLACIVL11TAAGSVGCSHFND12SG
   orf140-1        FLVALALRIAQGSATV9LTAAALMAPAVA10AAAGFTDWQLACIVL11TAAGSVGCSHFND12SG
20  orf140ng-1.pep  FWLVGRLLDM13DVPTTLK14TWTVNQT15LIAF16IGFALSALLFAIV
   orf140-1        FWLVGRLLDM13DVPTTLK14TWTVNQT15LIAL16IGFALSALLFAIV

```

Furthermore, ORF140ng-1 is homologous to an *E.coli* protein:

```

20  gi|882633 (U29579) ORF_o454 [Escherichia coli] >gi|1789097 (AE000358) o454;
   This 454 aa ORF is 34% identical (9 gaps) to 444 residues of an approx. 456 aa
   protein GNYP_ECOLI SW: F46832 [Escherichia coli] Length = 454
   Score = 210 bits (529), Expect = 1e-53
   Identities = 130/384 (33%), Positives = 194/384 (49%), Gaps = 19/384 (4%)
25  Query: 88  ETSGGAQSLADALIRMFGEKRAPFAPGVASLI1FGFPIFFDAGLIVMLPIV2FATARRMKQD 147
   E SGA+SLA+ R G+KR A +A+ G P+FFD G I++ PI++ A+ K
   Sbjct: 80  EHSGGAESLANYFSRKLGD1KRTIAALT2LA3AF4FLGIPV5FDVGFIILAPIYGF6AKVAKIS 139
30  Query: 148  VLPFALASVGAFSVMHVFLPPHPGPIAASEFYGANIGQVLI1LG2LTAFITWYFSGYMLGK 207
   L F L G +HV +PPHPGP+AA+ A+IG + I+G+ + I GY K
   Sbjct: 140  PLKFGLEPVAGIMLT1VHVAVPPHPG2PVAAAGL3HADIGWLT4IIGIAIS-IPGVVGYFAAK 198
35  Query: 208  VLGRAIHVPVPELL-----SGGTQSD1DPKPEKAGT2VVAVMLIPMLLI3FLNTGV 257
   ++ + + + E+L G T+ SD P A V +++++IP+ +I T
   Sbjct: 199  I1NKRQYAMSVEVLE2QMQLAPASEEGAT3KLSDKINPPGVA-LVTS4LVIVFIATIMAGT-- 255
40  Query: 258  SALISEKLVSADETWVQ1TAKMIGSTPXXXXXXXXXXXXXXXXX2GRKRGESGSTLEKTVDGALA 317
   +S L+ + T ++IGS +RG S +AL
   Sbjct: 256  ---VSATLMPPSH1LLGT2LQLIGSFMV3ALMIALV4LAFL5LLALRRGWSLQ6TS7DMIGSALP 312
45  Query: 318  PACSVILITGAGGMFGGVL1RASGIGKALADSMADLGI2PVLLGCFLVALALRIAQGSXXXX 377
   A VIL+TGAGG+FG VL SG+GKALA+ + +P+L F+++LALR +QGS
   Sbjct: 313  TAAVVLIVT1GAGGVFGKVL2VESGVGKALANMLQ3MDI4PLLPAAFI5ISLALRASQGS--AT 370
50  Query: 378  XXXXXXXXXXXXXXXX1GFTDWQLACIVLATAAGSVGCSHFNDSGFWLVGRLLDM2DVPTTLK 437
   G Q + LA G +G SH NDSGFV+V +L +V LK
   Sbjct: 371  VAILTTGGLSEAV1MLNP2IQCVL3TLAACFGGLGASHINDSGFWIV4TRYK5GLSVADGLK 430
   Query: 438  TWTVNQT1LIAF2IGFALSALLFAIV 461
   TWTV T++ F GF ++ ++A++
   Sbjct: 431  TWTVLT1ITILGFT2GPLIT3WCVNAVI 454

```

Based on this analysis, including the identification of the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

55 Example 71

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 593>:

```

1  ..GATTTCGGCA TATCGCCCGT GTATCTTTGG GTTGCCGCCG CGTTCAAACA
51 TTTGCTGCTCG CCGTGGGCTG CCGACTCATA CGATGTCGCA CGCTTTCGAC

```

101 GCGTATTTT TSCCGTTATC GGA CTGACTT CCTGCGGCTT TGCCGGTTTC
 151 AACTTTTGG CGAGACACA CGGCGCAC. GTCGCTCTGA TTCTCATCGG
 201 CTGTATCGGG CTGATTCAG TTGCCATT CTCAACCCC GCTGCGCGCG
 251 CCTTTGCCGC CGCGGACTG GTGCTGCACG GTTATTCTTT GGCTCGCGCG
 301 CGSGTGATT CGGCTCTTT TCTGCTCGGT ACGGCTGGA CGGTGATGTC
 351 GTTGGCAGCA GCTTATCCGG CAGCATTTGC CCTGATGCTG CCCTTGCCCG
 401 TACTGATGTT TTTCCGTCGG ..

This corresponds to the amino acid sequence <SEQ ID 594; ORF141>:

1 ..DFGISPVYLN VAAAFKHLSS PWAADSYDVA RFAGVFFAVI GLTSCGFAGF
 51 NFLGRHHGRX VVLLILIGIG LIPVAHFLNP AAAFAAAGL VLHGYSLARR
 101 RVIAASFLLG TGWTLSLAA AYPAAAFALML PLPLVLMFFRP ..

Further work revealed the complete nucleotide sequence <SEQ ID 595>:

1 ATGCTGACCT ATACCCCGCC CGATGCGCGC CGGCCGCCA AAACCCACGA
 51 AAAGCCGTGG CTGCTGCTGT TGATGCGGTT TGCCGTGGTG TGGCCCGGG
 101 TGTTTTCCCA CGATTGTGG AATCCTGACG AACCTCGGT CTATACCCGC
 151 GTCGAAGCAC TGGCAGGCAG CCCACCCCC TTGGTTGCCC ATCTGTTGCG
 201 TCAAACCGAT TTCCGCATAC CGCCCGTGTA TCTTTGGTT GCGCCGCGGT
 251 TCAAACTATT GTGTGTCGCG TGGGCTGCGG ACTCATACGA TGGCGCAGCG
 301 TTTTGAGGCG GATTTTTTCG GCTTATCGGA CTGACTCTCT GCGGCTTCC
 351 CGGTTTCAAC TTTTGGGCGA GACACACGCG GCGGCTGCTC GCTCTGATTC
 401 TCATCGGCTG TATCGGCTG ATTCCAGTGT CCGATTCTCT CAACCCCGCT
 451 CGCGCGCCT TTGCGCCGCG CGGACTGTGT CTGACGGTT ATTCTTTGGC
 501 TCGCGCGCGC GTGATTGCGC CCTCTTTCT GTCGCTACG GGCTGGAAGC
 551 TGATGTCGTT GGCAGCAGCT TATCGGCGAG CATTTGCCCT GATGCTGCCC
 601 TTGCCCCGAC TGATGTTTTT CCGTCCGTGG CAAGCAGCG GTTTGATGTT
 651 GACGGCAGTC GCTCTACCTT CCTTTGCCCT GCGGCTTATG ACCGTTTACC
 701 CGCTGCTCTT GGCAAAACG CAGCCCGCGC TGTTGCGCA ATGGCTCGAC
 751 TATCACGTTT TCGGTACGTT CCGCGCGGCT CGGCACGTT AGACGGCAT
 801 CAGTTTGTTT TACTATCTGA AAAACCTGCT TTGTTTGCA TTGCCCGCGC
 851 TGCCGCTGGC GGTTTGGAAG GTTTGCCGA GCGGCTGTT TTGCACCGAC
 901 TGGGGGATTT TGGGCGTCTG CTGGATGCTT GCGGTTTTGG TGCTGCTTGC
 951 CGTCAATCCG CAGCGTTTTT AGGATAACCT GCTCGGCTG CTTCCGCGCG
 1001 TTGCCCTGTT CCGCGCGGCG CAACCTGACA GCCTGAGGCG CCGCGCGCGG
 1051 GCGTTTGTC ACTGTTTGGC CATTATGGG TTCCGACTGT TTGCGCTGTT
 1101 CCTGTGACG GCTTTTTTCG CATTGAATTA CGGTCGCCC GCGAAGCTG
 1151 CGGACGCGC CGCTATTTC AGCCGTATT ATGTTCCGTA TATCGATCCC
 1201 ATCCGATGG CGGTGCGGCT ACTGTTACA CCTTTGTCG TGGGGCGAT
 1251 TACCCGAAA AACATACGCG GCAGCGAGCG GGTTACCAAC TGGCGGCGAT
 1301 CGGTTACCCT GACCTGGGCT TTGCTGATGA CGCTGTTCT GCGCTGGCTG
 1351 GACGCGCGA AAAGCCACGC GCGGCTGCT CGGAGTATGG AGGCATCGCT
 1401 TTCCCCGAA TTGAACGGG AGCTTTCAGA CGCATCGAG TGTATCGCA
 1451 TAGGCGCGG CGACCTGCAC ACGCGGATT TTTGACGCA GTACGCGACA
 1501 TTGCGCGACC CGGTGCGGCA TGTACAAATG CGCTACCGCA TCGCTCTCCT
 1551 GCCCAAAAT GCGGATGCGC CGCAAGGCTG CACAGCGGTT TGGCAGGCT
 1601 CGGCTCCGCG CAACAAAGAC AGTAAGTTGC GACTGATAC GAAATCCGGG
 1651 GAAATATAT AA

This corresponds to the amino acid sequence <SEQ ID 596; ORF141-1>:

1 MLTYTPFPDAR PPAKTHEKFW LLLIMAFWL WPGVFSHDLW NPDEPAVYTA
 51 VEALAGSTPT LVAHLFGQTD FGIPVYLVW AAFAKHLSP WAADSYDAR
 101 FGVVFAVIG LTSCGAFGN FLGRHGRSV VLLILIGIG LIPVAHFLNP
 151 AAFAAAAGLV VLHGYSLARR VTAASFLLG GTWTLSLAA YPAFAALML
 201 LPVLMFFRPW QSRRLMTAV ASLAFALPLM TYPYLLAKT QPALFAQWLD
 251 YHVFGTFGV RHVQTAFSLF YLKNLLWFA LPAFLPVMVT VCRTRLFSTD
 301 WGLIGVYVML AVLVLLAVNP QRFDNLVWL LPPIALFGAA QDLSLRGAA
 351 AFVNWFGIMA FGLFAVFLWT GFFAMNWWP AKLAERAYF SPYVVPDIDP
 401 IPMAVAVLET PLWLWATR K NIRGQAVTN WAAGVTLTWA LLMTLFLPLW
 451 DAAKSHAPVV RSMASLSPE LKRELSGIE CIGTGGDLH TRIVWTQYGT
 501 LPHRVGDVQC RYRIVLLPN ADAFQGWQTV WQGARPRNKD SKFALRKIG
 551 ENI*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF141 shows 95.0% identity over a 140aa overlap with an ORF (ORF141a) from strain A of *N. meningitidis*:

```

5      orf141.pep      DFGISPVYLWVA10AAAFKHL20SPWAADSYDVA30
      orf141a      WNPDEFAVYTAVEALAGSPPTPLVAHLFGQIDFGIPFVYLWVA40AAAFKHL50SPWAADPYDAA60
      orf141.pep      RFAGVFFVAVIGL40TSCGFAGNFELGRHHGRXV50VLILGICIGLI60PEVAHFLNFAAA70FAAAGL80
      orf141a      RFAGVFFVAVIGL40TSCGFAGNFELGRHHGRS50VVLILGICIGLI60PTVFH70LNFAAA80FAAAGL90
10     orf141.pep      VLHGYSLARRRVIAASFL100LGCTGWTLM110SLAAAYPAAFAL120MLEPLVLMFERP130
      orf141a      VLHGYSLARRRVIAASFL100LGCTGWTLM110SLAAAYPAAFAL120MLEPLVLMFFERQWSRRIM130MLTA140
15     orf141.pep      VASLAFALPILMTVYPLLAKTOPALFQWLDDHVF160GTFGGV170VRHIQTAFS180IFYLLKNLWF190
      orf141a      VASLAFALPILMTVYPLLAKTOPALFQWLDDHVF160GTFGGV170VRHIQTAFS180IFYLLKNLWF190
20     orf141a      VASLAFALPILMTVYPLLAKTOPALFQWLDDHVF160GTFGGV170VRHIQTAFS180IFYLLKNLWF190

```

The complete length ORF141a nucleotide sequence <SEQ ID 597> is:

25	1	ATGCTCGACCT	ATACCCGCCGC	GGATGCCCGC	CGCCCGCCCA	AAACCCACGA
		AAGAGCGTGG	TCGTGCTGTG	GGATGGCGTT	TGCGTGGTGT	TGGCCCGCGA
	101	TGTTTTCOCA	CGATTGTGTG	AATCCCTGAC	AACTCTGCCG	CTATACCGCG
	151	GTGCAAGCAC	TGGCAGGCAG	CCACCCACTC	TGTGGTGGCT	ATCTGTTTGG
30	201	TCAAACTCAT	TTCGGCATAT	CGGCCGTGTA	CTTTTGGGTT	CGCCGCGCGT
		TCAAACATTT	GCTGTGCGCG	TGGGCTGGCG	ACCGGTATGA	TGCCGCACTC
	301	TTTGCSCGCG	TTTTTTTCGC	GCTTGTCCGA	CTGACTTCCT	CGGGCTTTGC
	351	CGGTTTCAAC	TTTTTGGGCA	GACACACAGA	CGGCAGCTCG	GTCTGATTTC
	401	TCATCGGGCT	TATCGGGCTG	CTTCCGACGC	TACACTTTCT	CRAOCCCGCT
	451	CGCGCGCGCT	CGACGCGCGT	AGGACTGGTG	TGCGAGGTTT	ATTCTTTTGC
35	501	TGCGCGGGCG	GTGATTGCGC	CTCTCTTTCT	CGTTCGACG	GTTTGGAGCG
		TGATGTTCGT	TGGAGCAGCT	TATCCSGGCG	CATTGTCCT	GATGTCGCTG
	601	CTGCGCGGCG	TGATGTTTTT	CCGCTCGTGG	CAGACGACG	GTTCGATGCT
	651	GACGCGASTC	GCGCTGCTCG	CGTTTCTGCG	CGGCTGATCG	ACCGTTTACG
	701	CGCTGCTCTG	CGCAAAACAG	CAGCCGCGCG	TGTTTCCGCA	ATGGCTCTAC
40	751	GATCAGCTTT	TGCTGTACGTT	CGCGCGCGCT	CGGCACATTT	AGACGGCAAT
	801	CAGTGTGTTT	TACACTCTGA	AAACACTGCT	TGCTTTTGA	TTGCGTCGCG
	851	TGCGCGCTGC	GTTTGGAGCG	TGTTGCCGCA	CGCGCGCTCT	TTCGACCGAC
	901	TGGGGGATTT	TGGGGGTCGT	CTGGATGCTT	CGCGTTTGG	TGCTGTCTGC
	951	CGTCAACTCG	CAGCGTTTTC	AGGATAAACC	CGCTGGAGTG	CTTCCGCGCG
45	1001	TTCGCTCTGT	CGCGCGGCGC	CAGTACGGCA	GCGTGGAGCG	CGGCGCGGCG
	1051	CGGTTTGTCA	ACTGGTTTGC	CATTATGGCC	TTCCGACTCT	TTGCCGCTGT
	1101	CTCTGGAGAC	GGCTTTTTCG	CCATGAATGA	CGGCTGCCCC	GCCAAGCTGT
	1151	CCGAACGCGC	CGCATTATTC	AGGCCGTAAT	TCGTTCTCTGA	TATGATCGAT
	1201	ATTCCGATGT	CGGTTGCGGT	ACTGTTGCCA	ACGTTTGGSC	TGTGGGCACT
50	1251	TGCCCGGAAA	AACATACCGG	CGAGCGAGCG	CGGTACCAAC	TGGCGGCGAC
	1301	GAGTTTACCT	CTAGTGGGCT	TGCTGTATGA	CGGTATCTCT	GCCTGTACCT
	1351	GAGCGACAGA	AGAGCGTCTG	CGGATGACCA	CGGATGACCA	AGGATGACCA
	1401	TTCCCGCGAA	TTAAACCGCG	AGCTTTTCTA	CGGATCGAG	TGTATCGACA
	1451	TAGCGCGGCG	CGACCTACAC	ACGCGGATGT	TTTGGACGCA	GTAGCGGAGC
55	1501	TTCGCCGACC	CGGTCGCGCA	TGTACAATGT	CGCATACGCA	TGCTCGGCTT
	1551	GGCCCAAAAC	GGGTGATGCG	CGCAAGGCTG	CAGACAGCTC	TGSCAGGGTGT
	1601	CGCGCGCGCG	CACAACAGAC	AGTAAGTTTG	CAGTGTATCG	GAARAACCGG
	1651	GAAATATATAT	TAAAAACAAC	AGATTGTA		

This encodes a protein having amino acid sequence <SEQ ID 598>:

60 1 MLTYTPPDPAR PPAKTHEKPW LLLMAFAWL WPGVFSHDLW NPDEPAVYTA
51 VEALAGSPTP LVAHLFGQID FGIPPVYLWV AAAPKHLSP WAADPYDAAR
101 FAGVFFFAVVG LTSCGFAGEN FLGRHHGRSV VLILIGICGL IPTVHFLNPA

151 AAFAAAGLV LHGYSLARRR VIAASFLGT GWTLMSLAAA YPAAFALMLP
 201 LPVLMFFRPW QSRRLMLTAV ASLAFALPMT TVYPILLAKT QPALFAQWLD
 251 DHVFGTGGV RHQIQTAFSLF YYLKNLLWFA LPALPLAVMT VCRTRLFSTD
 301 WGILGVVWML AVLVLAVNP QRFQDNLVWL LPPLALFGAA QLSLRRGAA
 351 AFVNWFGIMA FGLFAVFLWT GFFAMNYGWP AKLAERAAFY SPYYVEDIDP
 401 IPMAVAVLFT PLWLWAIKTRK NIRGRQAVTN WAAGVTLWA LMTLFLPWL
 451 DAAKSHAPV RSMESASPE LKRELSDGIE CIDIGGGDLH TRIVWTQYGT
 501 LPHRVGDVQC RYRIVRLPQN ADAPOGQWTV WQGARPRNKD SKFALIRKGT
 551 ENILKTTD*

10 ORF141a and ORF141-1 show 98.2% identity in 553 aa overlap:

15	orf141a.pep	MLTYTPPDARPPAKTHEKPWLLLLMAFAWLWPGVFSHDLWNDEPAVYTAVEALAGSPTP
	orf141-1	MLTYTPPDARPPAKTHEKPWLLLLMAFAWLWPGVFSHDLWNDEPAVYTAVEALAGSPTP
20	orf141a.pep	LVAHLFGQIDFGIPPVYLVWAAAFKHLLSPWAADPYDAARFAGVFFAVVGLTSCGFAGFN
	orf141-1	LVAHLFGQIDFGIPPVYLVWAAAFKHLLSPWAADSYDAARFAGVFFAVIGLTSOGFAGFN
25	orf141a.pep	FLGRHHGRSVVLLIGICIGLIPTVHFLNPAAAFAAAGLVLHGYSLARRRVIAASFLIGT
	orf141-1	FLGRHHGRSVVLLIGICIGLIPVAHFLNPAAAFAAAGLVLHGYSLARRRVIAASFLIGT
30	orf141a.pep	GWTLMSLAAAYPAAFALMLPLPVLMMFFRPWQSRRLMLTAVASLAFALPMTVYPILLAKT
	orf141-1	GWTLMSLAAAYPAAFALMLPLPVLMMFFRPWQSRRLMLTAVASLAFALPMTVYPILLAKT
35	orf141a.pep	QPALFAQWLDHVFVFGTFGGVRHIQTAFSLFYLLKNLLWFALPALPLAVMTVCRTRLFSTD
	orf141-1	QPALFAQWLDYHVFVFGTFGGVRHVQTAFSLFYLLKNLLWFALPALPLAVMTVCRTRLFSTD
40	orf141a.pep	WGILGVVWMLAVLVLAVNPQRFQDNLVWLPPALFGAAQLSLRRGAAAFVNWFGIMA
	orf141-1	WGILGVVWMLAVLVLAVNPQRFQDNLVWLPPALFGAAQLSLRRGAAAFVNWFGIMA
45	orf141a.pep	FGLFAVFLWTGFFAMNYGWPAKLAERAAFYSPYYVPDIDPIPMNAVAVLFTPLWLWAIKTRK
	orf141-1	FGLFAVFLWTGFFAMNYGWPAKLAERAAFYSPYYVPDIDPIPMNAVAVLFTPLWLWAIKTRK
50	orf141a.pep	NIRGRQAVTNWAAGVTLTWALLMTLFLPWLDAAKSHAPVVRSMESASPELKRELSDGIE
	orf141-1	NIRGRQAVTNWAAGVTLTWALLMTLFLPWLDAAKSHAPVVRSMESASPELKRELSDGIE
	orf141a.pep	CIDIGGGDLHTRIVWTQYGTLPHRVGDVQCERYRIVRLPQNADAPQGQWTVWQGARPRNKD
	orf141-1	CIDIGGGDLHTRIVWTQYGTLPHRVGDVQCERYRIVRLPQNADAPQGQWTVWQGARPRNKD
	orf141a.pep	SKFALIRKGTGENI
	orf141-1	SKFALIRKGTGENI

Homology with a predicted ORF from *N.gonorrhoeae*

ORF141 shows 95% identity over a 140aa overlap with a predicted ORF (ORF141ng) from *N.gonorrhoeae*:

55	orf141.pep	DFGISPVYLVWAAAFKHLLSPWAADSYDVA	30
	orf141ng	WNPAEPVYTAVEALAGSPTPLVAHLFGQIDFGIPPVYLVWAAAFKHLLSPWAADSYDVA	126
	orf141.pep	RFAGVFFAVIGLTSOGFAGFNFLGRHHGRXVLLIGICIGLIPVAHFLNPAAAFAAAGL	90
60	orf141ng	RFAGVFFAVIGLTSOGFAGFNFLGRHHGRSVVLIHIGICIGLIPVAHFLNPAAAFAAAGL	186
	orf141.pep	VLHGYSLARRRVIAASFLIGTGWTLMSLAAAYPAAFALMLPLPVLMMFFRP	140
	orf141ng	VLHGYSLARRRVIAASFLIGTGWTLMSLAAAYPAAFALMLPLPVLMMFFRPWQSRRLMLTA	246

An ORF141ng nucleotide sequence <SEQ ID 599> was predicted to encode a protein having amino acid sequence <SEQ ID 600>:

```

1  MPSEAVSARP  LCEYLLHLAI  RFLLTLMLT  YTPFDARPEA  KTHEKFWLLL
5  1  LMAFAWLWPG  VFSHDLWNPA  EPAVYTAVERA  LAGSPTPLVA  HLFQQTDEGI
101 1  FVYVLWVAAA  FKHLSPWAA  HPYDAARFAG  VFFAVIGLTS  CGFAGFNLFG
151 1  RHHGRSVLLI  HIGCIGLIPV  AHFFNPAAAA  FAAAGVLVHG  YSLARRVIA
201 1  ASFLLGTGWT  LMSLAAAYPA  AFALMLPLPV  LMFFRPQWSR  RLMLTAVASL
251 1  AFALPLMTVY  PLLLAKTQPA  LFAQWLNVHV  FGTGGVRHRI  QRAFSLFHYL
301 1  KNLLWFAPFG  LPLAVTWCR  TRLFSTDWGI  LGIVWMLAVL  VLIAPNPQRF
10 301 QDNLVWLPP  LALFGAAQLD  SLRRGAAAFV  NWFGIMAFGL  FAVFLWTGFF
401 1  AMNYGWPAKL  AERAAVFSFY  YVPDIDIPM  AVAVLFTFLW  LWAITRKNIR
451 1  GRQAVTNWAA  GVTLTWALLM  TLFLEWLDAA  KSHAPVVRSM  EASFSPELKR
501 1  ELSDGIECIG  IGGGDLHTRI  VWTQYGTLPF  RVGDVRCRYR  IVRLPQNADA
551 1  FQGWQTVWQG  ARPRNKDSKF  ALIRKIGENI  LKTTD*

```

Further work revealed the following gonococcal DNA sequence <SEQ ID 601>:

```

1  ATGCTGACCT  ATACCCCGCC  CGATGCCGCG  CCGCCGCCCA  AAACCCACGA
5  1  AAAACCGCTGG  CTGCTGCTGT  TGAATGGCGT  TGCCCTGGCT  TGGCCCGCGG
101 1  TGTITTTCCCA  CGATTGTGTG  AATCCTGCCG  AACCTGCCCT  CTATACCCGC
151 1  GTGCAAGCAC  TGGCAGGCAG  CCCCACCCCC  TTGGTGGCCC  ATCTGTTGGG
201 1  TCAAAACGAT  TTGCGCATAC  CGCCCGTGTA  TCTTTGGGTT  GCCCGCGCAT
251 1  TCAAAACATT  GCTGTGCGCG  TGGGCAGCCG  ACCCGTAGTA  TGGCCGACGC
301 1  TTTGCAAGCG  TAITTTTTCG  CGTTATCGGA  CTGACTTCTT  CGCGCTTTGC
351 1  CGGTTTCAAC  TTTTGTGGCA  GACACCAACG  CGGCAGCGTT  GTTTTAATCC
401 1  ATATCGGCTG  TATCGGCGTG  ATTCCGGTTG  CCATTTCTCT  CAATCCCGcc
451 1  gccgcgcgcct  tTGC CGCGCG  CGGA CTGGTG  TGCacggcct  actcgctggc
501 1  ACGCCGCGCG  GTGATTCggc  cctctTtcc  GTCCTGGTAC  GGTGTGACGT
551 1  TGAATGTCGT  GCGCGGCAGT  TATCGCGGCG  GTTGTGGCGT  GATGCTGGCC
601 1  CIGCCCGTGC  TGAITTTTTT  CCGTCCGCTG  CAAAGCAGCG  GTTTGATGTT
651 1  GACCGGACCT  GCGCTGCCCT  CCGCTTATCG  ACCGCTTATC  ACCTGTTTAC
701 1  GCTGCTCTCT  gGCAAAACG  CAGCCCGCGT  TGTTTGCGCA  ATGGCTCAAC
751 1  TATCACGTTT  TCGGTACGTT  cggcgcgcgt  cgggttcgca  ccgcgcgcgt
801 1  Cagtttgttt  cactatctga  AAaatctcgt  ttggttcgca  ccgcgcgcgt
851 1  TGCCGCTGGC  GGTTTGAGCG  GTTTGCGCGA  CAGCGCTGTT  TTCAGCCGAC
901 1  TGGGGGATTT  TGGGCATTGT  CTGGATGCTT  GCGGTTTGGT  TGCTGCTCGC
951 1  CTTTAAATCC  CAGCGTTTTC  AAGACAACT  GGTCTGGCTG  CTGCGCGCGC
1001 1  TTGCCCTGTT  CGCGCGCGCG  CAACTGGACA  CGCTGAGGCG  CGGCGCGCGC
1051 1  GCTTTTGTC  ACTGTTTCGG  CATATGCGCG  TTCGGGCTGT  TTGCCGTGTT
1101 1  CCGTGGACG  GGCTTTTTCG  CAGTAATTA  CGGCTGGCCC  GCCAAGCTTG
1151 1  CCGAACGCGC  CGCTACTTCT  AGCCGTATT  ACGTTCCCGA  CATCGATCCC
1201 1  ATTCCGATGA  CGGTTGCGGT  ACTGTTTACA  CGCTTGTGCG  TTTGGGCGAT
1251 1  TACC CGGARA  AACTATACCG  CGAGGCAGCG  GGTTACACAA  TGGGCGCGCAG
1301 1  GCGTTACCT  GACTCGGCT  TGTCTGATGA  CGCTGTTCCT  CGCGTGGCTG
1351 1  GAGCGCGCGA  AAGCCACGCG  CAGCGTCCT  CGGAGTATCG  AGCGATCGTT
1401 1  TTCCCGCGAA  TTAACACGG  AGCTTTCAGA  CGCATTCGCA  TGTATCGCA
1451 1  TAGGCGGCGG  CGACTGCAC  ACCGCGGAT  TTTGACACCA  GTACGGCACA
1501 1  TTGCGCGACC  GCGTCCGCG  TGTCCGTGCG  CGCTACGCTA  TCGTCCGCTG
1551 1  GCCCCAAAC  CGGATGCGC  CGCAAGGCT  CGAGACGCT  TGGCAGGCTG
1601 1  CGCGCCCGCG  CAACAAAGC  AGTAAGTTTG  CACTGATACG  GAAATTCGG
1651 1  GAAATATAT  TAAACACAC  AGATTGA

```

This corresponds to the amino acid sequence <SEQ ID 602; ORF141ng-1>:

```

1  MLTYTPPDAR  PPAKTHEKFW  LLLLMFAWL  WPGVFSHDLW  NPAEPAVYTA
5  1  VEALAGSPTE  LVAHLFGQTD  FGIPPVYLMV  AAAPKHLISP  WAADPYDAAR
101 1  FAVVFVAVIG  LITSCGFAGN  FLGRHGRGSV  VLTHIGCIGL  IFAHFLNPA
151 1  AAFAAAGLY  LUGYSLARR  VLAASFLTGT  GWTLMSLAA  YPAAFALMLP
201 1  LPLMFEPFW  QSRNMLTAV  ASLAFALPL  TGVLLTAKT  QPALFAQWL
251 1  YHVFETFGV  RHQRAFSLF  HYLENLWLF  PGLPLAVMT  VCTRLFSTD
301 1  WGILGIVWML  AVLVLLAENF  QRQDNVLVL  LPLPALFGAR  QLDSIRGARA
351 1  AFVNWFGIMA  FGLFAVFLMT  GTFAMNYGWF  AKLAERAAFY  SPYYVPDIDP
401 1  IMAVAVLFT  PLWLWATR  NIRGRQAVTN  WAAGVTLTWA  LLMTLFLPWL
451 1  DAAKSHAPV  RSMEASFSP  LKRELSDGIE  CTIGGGDLH  TRIVWTQYGT
501 1  LPHRVGDVRC  RYRIVRLPQN  ADAPGQWTV  WQGARPRNKD  SKFALIRKIG
551 1  ENILKTTD*

```

ORF141ng-1 and ORF141-1 show 97.5% identity in 553 aa overlap:

```

5  orf141ng-1.pep  MLTYTPPDARPPAKTHEKFWLLLLMAFAWLPVGFSDHNLNPAEFAVYTAVEALAGSPTF
   orf141-1       MLTYTPPDARPPAKTHEKFWLLLLMAFAWLPVGFSDHNLNDEFAVYTAVEALAGSPTF
10  orf141ng-1.pep  LV AHLFGQTDFGIPPVYLWVAAAFKHL LSPWAADPYDAARFAGVFVAVIGLTS CGFAGFN
   orf141-1       LV AHLFGQTDFGIPPVYLWVAAAFKHL LSPWAADS YDAARFAGVFVAVIGLTS CGFAGFN
15  orf141ng-1.pep  FLGRHRCRSVVLILIGICIGLIPVAHFLNPAAAAFAGLV LHGYSIARRRVIAASFL LGT
   orf141-1       FLGRHRCRSVVLILIGICIGLIPVAHFLNPAAAAFAGLV LHGYSIARRRVIAASFL LGT
20  orf141ng-1.pep  GWTIMSLAAAYPAAFALMLPLPVLMMFFRPWQSRRLMLTAVASLAFALPLMTVYPLLLAKT
   orf141-1       GWTIMSLAAAYPAAFALMLPLPVLMMFFRPWQSRRLMLTAVASLAFALPLMTVYPLLLAKT
25  orf141ng-1.pep  QPALFAQWLNYHVEGTFGGVRHIQRAFSLFHYLKNLLWFAPGGLPLAVWTVCRTRLFSTD
   orf141-1       QPALFAQWLNYHVEGTFGGVRHVQTAFSLFY LKNLLWFALPALPLAVWTVCRTRLFSTD
30  orf141ng-1.pep  WGILGVWMLAVLVLAVNPFQPDNLVWLLPPLALFGAAQLDSLRGAAAFVNWFGIMA
   orf141-1       WGILGVWMLAVLVLAVNPFQPDNLVWLLPPLALFGAAQLDSLRGAAAFVNWFGIMA
35  orf141ng-1.pep  EGLFAVFLMTGFFAMNYGWPAKLAERAAYFSPYVDPIDIPMAVAVLFTPLWLWAI TRK
   orf141-1       EGLFAVFLMTGFFAMNYGWPAKLAERAAYFSPYVDPIDIPMAVAVLFTPLWLWAI TRK
40  orf141ng-1.pep  NIRGRQAVTNWAAGVTLTWALMTLFLPWLDAAKSHAPVVRMEASFSPELKRELS DGE
   orf141-1       NIRGRQAVTNWAAGVTLTWALMTLFLPWLDAAKSHAPVVRMEASFSPELKRELS DGE
45  orf141ng-1.pep  CIGIGGGDLHTRIVWTQYGTLP HRVGDVQCRYRIVRLPQNADAPQGQWTVWQGARPRNKD
   orf141-1       CIGIGGGDLHTRIVWTQYGTLP HRVGDVQCRYRIVRLPQNADAPQGQWTVWQGARPRNKD
50  orf141ng-1.pep  SKFALIRKIGENILKTTDX
   orf141-1       SKFALIRKIGENIX

```

Based on the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 72

45 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 603>:

```

1  ..CBATCGGCCA AATGTTTATC GGGCCAACT CTAGTCGGCA CAGCAATTGG
51  GATACGCGGG CAGATAAAGC TTGGCGGCAA CTGCATTAC GATATATTTA
101 CCGGCGCGGC ATTGAAAAAG CCCGAATTTT TCCAATCAAG GAAATGGGCA
151 AGCGGTTTTC AGGTAGGCTA TACGTTTAA

```

50 This corresponds to the amino acid sequence <SEQ ID 604; ORF142>:

```

1  ..QSAKWLSGQT LVGTAIGIRG QIKLGNLHY DIFTGRALKK PEFFQSRKWA
51  SGFQVGYTF*

```

Further work revealed the complete nucleotide sequence <SEQ ID 605>:

```

55 1  ATGATAATT CGGTAAGTGA GGCACAGGA AAATACCAAG GAAATATCAC
   51 TTTCTCTGCC GACAATCCTT TGGGACTGAG TGATAGTTTC TATGTAATT

```

101 ATGGAAGTTC GATTGGCGGT ACGCCGATG AGGAAAGTTT TGACGGCCAT
 151 CGCAAGAAG GCGGATCAAA CAATTACGCC GTACATTATT CAGCCCTTTT
 201 CGTAAATGG ACATGGGCAT TCAATCACAA TGGCTACCGT TACCATCAGG
 251 CAGTTTCCGG ATTATCGGAA GTCTATGACT ATAATGGAAA AAGTTACAAAT
 301 ACTGATTTCG GCTTCAACCG CTTGTTGTAT CGTGATGCCA AACGCAAAAC
 351 CTATCTCGGT GTAAAACGTG GGATGAGGGA AACAAAAGT TACATTGATG
 401 ATGCGCAACT GACTGTACAA CGGCGTAAAA CGCGGGTTG GTTGCCAGAA
 451 CTTTCCCA CAAGAATATAT CGGTCGCA GTGCGCGCG CTTGAAGTAA
 501 ATATAAACCG GGCACCGGCA TGAAGATGCG TTAGGACATC
 551 CCTTTGGCGA AGGCACGTCA CGTATGAAA TTTGGACGCG ATCGGCTGAT
 601 GTAAATACTC CTTTTCAAAT CGGTAAACAG CATTGTGCGT ATGACACATC
 651 CGTTCATGCA CAATGGAACA AAACCCCGCT AACATCGCAA GACAAACTGG
 701 CTATCGGCGG ACACCAACCC GTACGTGGCT TCGACGGTGA AATGAGTTTG
 751 TCTGCGGAGC GGGGATGGTA TTGGCGCAAC GATTTGAGCT GGCAATTTAA
 801 ACCAGGCCAT CAGCTTTATC TTGGGCGTGA GTTAGGACAT GTTTCAGGAC
 851 AATCGCCCAA ATGTTTATCG GGCACCACTC TAGTCCGCAC AGCAATGGGG
 901 ATACGCGGGC AGATAAAGCT TGGCGCAAC CTGCATTACG ATATATTATC
 951 CGGCGCGGCA TTGAAAAGC CCGAATTTT CCAATCAAGG AAATGGGCAA
 1001 GCGGTTTTCA GGTAGGCTAT ACGTTTTTAA

This corresponds to the amino acid sequence <SEQ ID 606; ORF142-1>:

1 MDNSGSEATG KYQGNITFSA DNPGLSDMF YVNYGRSIGG TFDEESFDGH
 51 RKEGGSNNYA VHYSAFFKQW TWAFNHNRYR VSGVSLSE VYDYNKGSYN
 101 TDFGFNRLLY RDAKRTKTYL VKLWMRETKS YIDDAELTVQ RRKTAGWLAE
 151 LSHKEYIGRS TADFCLKYKR GTGMKDALRA PEEAFEGETS RMKIWTASAD
 201 VNTFPQIGKQ LFAYDTSVHA QWNKTPLTSS DKLAIGHHT VRFDGEMSL
 251 SAERGWYWRN DLSWQFKPHG QLYLGADVHG VSGQSFWMLS GQTLVGTALG
 301 TRGQIKLGGN LHYDIETGRA LKKEFFEQSR KWASGFQVGY TE*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.gonorrhoeae*

ORF142 shows 88.1% identity over a 59aa overlap with a predicted ORF (ORF142ng) from *N.gonorrhoeae*:

orf142.pep		QSAKWLSGQTLVGTALGIRGQIKLGGNLYH	30
orf142ng	RGWYWRNDSLWQFKPHGQLYLGDVGHVSGQSAKWLSGQTLGATGIRGQIKLGGNLYH		313
orf142.pep	DIFTGRALKKPEFFQSRKWASGFQVGYTF	59	
orf142ng	DIFTGRALKKPEYFQTKKWVTFQVGYSF	342	

The complete length ORF142ng nucleotide sequence <SEQ ID 607> is:

1 ATGGATAATT CGGGTAGTGA GGCACAGGA AAATACCAAG GAAATATCAC
 51 TTTCTCTGCC GACAATCCTT TTGGACTGAG TGATATGTTT TATGTAATTT
 101 ATGGAAGTTC AATTGGCGGT ACGCCGATG AGGAAAATTT TGACGGCCAT
 151 CGCAAGAAG GCGGATCAAA CAATTACGCC GTACATTATT CAGCCCTTTT
 201 CGTAAATGG ACATGGGCAT TCAATCACAA TGGCTACCGT TACCATCAGG
 251 CAGTTTCCGG ATTATCGGAA GTCTATGACT ATAATGGAAA AAGTTACAAAT
 301 ACTGATTTCG GCTTCAACCG CTTGTTGTAT CGTGATGCCA AACGCAAAAC
 351 CTATCTCAGT GTAAAACGTG GGACGAGGGA AACAAAAGT TACATTGATG
 401 ATGCGCAACT GACTGTACAA CGGCGTAAAA CCACAGGTTG GTTGCCAGAA
 451 CTTTCCCA CAAGAATATAT CGGTCGCA GTGCGCGCG CTTGAAGTAA
 501 ATATAAACCG GGCACCGGCA TGAAGATGCG TTAGGACATC
 551 CCTTTGGCGA AGGCACGTCA CGTATGAAA TTTGGACGCG ATCGGCTGAT
 601 GTAAATACTC CTTTTCAAAT CGGTAAACAG CATTGTGCGT ATGACACATC
 651 CGTTCATGCA CAATGGAACA AAACCCCGCT AACATCGCAA GACAAACTGG
 701 CTATCGGCGG ACACCAACCC GTACGTGGCT TCGACGGTGA AATGAGTTTG
 751 CTTGCGGAGC GGGGATGGTA TTGGCGCAAC GATTTGAGCT GGCAATTTAA
 801 ACCAGGCCAT CAGCTTTATC TTGGGCGTGA GTTAGGACAT GTTTCAGGAC
 851 AATCGCCCAA ATGTTTATCG GGCACCACTC TAGTCCGCAC AGCAATGGGG
 901 ATACGCGGGC AGATAAAGCT TGGCGCAAC CTGCATTACG ATATATTATC
 951 CGGCGCGGCA TTGAAAAGC CCGAATATTT TCGACAGGAG AAATGGGTA

1001 CGGGGTTTCA GGTGGGTTAT TCGTTTGA

This encodes a protein having amino acid sequence <SEQ ID 608>:

1 MDNSGSEATG KYQGNIITFSA DNPFGLSDMF VVNYGRSIGG TPDEENFDGH
 51 RKEGGSNNYA VHYSAFFGKW TWAFNHNGYR YHQAVSGLSE VYDINGKSYN
 101 TDFGFNRLLY RDAKRRTYLS VKLWTRRETS YIDDAELTVQ RRKTTGWLAE
 151 LSHKGYIGRS TADFKLKXKH GTCMKDALRA PEEAFGEGETS RRMKIWPASAD
 201 VNTDFPIGKQ LEAYDTSVHA QNNKTPLT SQ DKLAIGGHT VRGFDGEMSL
 251 PAERGWYWRN DLSWQFKPGH QLYLGADVGH VSGQSAKWL SQTLAGTAIG
 301 IRGQIKLGGN LHYDIFTGRA LKKPEYFQTK KWTGFQVGY SF*

The underlined sequence (aromatic-Xaa-aromatic amino acid motif) is usually found at the C-terminal end of outer membrane proteins.

ORF142ng and ORF142-1 show 95.6% identity over 342aa overlap:

orf142-1.pep MDNSGSEATGKYQGNIITFSADNPLGLSDMFVYVNYGRSIGGTPDEESFDGHRKEGGSNNYA
 orf142ng-1 MDNSGSEATGKYQGNIITFSADNPLGLSDMFVYVNYGRSIGGTPDEENFDGHRKEGGSNNYA
 orf142-1.pep VHYSAFFGKWTWAFNHNGYRYHQAVSGLSEVYDINGKSYNTDFGFNRLLYRDAKRRTYLSG
 orf142ng-1 VHYSAFFGKWTWAFNHNGYRYHQAVSGLSEVYDINGKSYNTDFGFNRLLYRDAKRRTYLS
 orf142-1.pep VKLWMRETSYIDDAELTVQRRKTAGWLAE LSHKEYIGRSTADFKLYKRGTMKDALRA
 orf142ng-1 VKLWMRETSYIDDAELTVQRRKTAGWLAE LSHKEYIGRSTADFKLYKRGTMKDALRA
 orf142-1.pep PEEAFGEGETSRMKIWTASADVNTFPQIGKQLFAYDTSVHAQNNKTPLT SQDKLAIGGHT
 orf142ng-1 PEEAFGEGETSRMKIWTASADVNTFPQIGKQLFAYDTSVHAQNNKTPLT SQDKLAIGGHT
 orf142-1.pep VRGFDGEMSLAERGWYWRN DLSWQFKPGHQLYL GADVGHVSGQSAKWL SGT LVGTAIG
 orf142ng-1 VRGFDGEMSLAERGWYWRN DLSWQFKPGHQLYL GADVGHVSGQSAKWL SGT LVGTAIG
 orf142-1.pep IRGQIKLGGN LHYDIFTGRALKKPEFFQSRKWASGFQVGYTF
 orf142ng-1 IRGQIKLGGN LHYDIFTGRALKKPEYFQTKKWTGFQVGYSF

In addition, ORF142ng is homologous to the HecB protein of *E. chrysanthemi*:

gi|1772622 (L39897) HecB [Erwinia chrysanthemi] Length = 558
 Score = 119 bits (295), Expect = 3e-26
 Identities = 88/346 (25%), Positives = 151/346 (43%), Gaps = 22/346 (6%)
 Query: 2 DNSGSEATGKYQGNIITFSADNPLGLSDMFVYVNYGRSIGGTPDEENFDGHRKEGGSNNYAV 61
 Score = 119 bits (295), Expect = 3e-26
 Identities = 88/346 (25%), Positives = 151/346 (43%), Gaps = 22/346 (6%)
 Sbjct: 230 DNSGQKSTGEEQLNGSLALDNVFLGADQWFI SAGHS---SRFATSDAESLQAG----- 280
 Query: 62 HYSAPFGKWTWAFNHNGYRYHQAVSGLSEVYDINGKSYNTDFGFNRLLYRDAKRRTYLSV 121
 Sbjct: 281 -FSMPGYVWNLGYNYSQSRKYNTFINRDFWHSTGDSDFTRFSLSRVFRDGTMTKTAIAG 339
 Query: 122 KLVWRETSYIDDAELTVQRRKTAGWLAE LSHKEYIGRSTADFKLYKRGTMKDALRAP 181
 Sbjct: 340 TFSQRTGNNYINGSLPLSSSRKLSVSLGVNHSQKLWGLGATFNPTNYNRGVRWLGSSETDT 399
 Query: 182 EEAFFGEGETSRMKIWTASADVNTFPQIGKQLFAYDTSVHAQNNKTPLT SQDKLAIGGHTV 241
 Sbjct: 400 DKSADFPRAEFNKWTLASASYHPV---TDSITLGLSYGQYSARALYGSQELTLGGESSI 456
 Query: 242 RGF D GEMSLAERGWYWRN DLSWQFKP----GHQLYLGA-DVGHVSGQSAKWL SGT LAG 296
 Sbjct: 457 RGF-REQYTSNGR GAYWRNELN WQAWQLPVLGNVTFMAAVDGGHLN HKQDNSTASALWG 515

Query: 297 TAIGIRGQIKLGGNLYHYDIFTGRALKKPEYFQTKKVVTFQGVGSF 342
 A+G + L + G + P + Q V G++VG SF
 Sbjct: 516 GAVGMTVASRW---LSQQVTVGWPIISYPALQPDIMVVGYRVGLSF 558

On the basis of this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 73

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 609>:

```

1 ATGCGGACGA AATGGTCAGC AGTGAGAAGC TGCTTACTTG GcCGGACACC
51 GCGGACATCG ATACCGCTTT GAACCTGTG TACCGTTTGC AAAAACTCGA
101 ATTCTCTAT GCGGATGAAA ACGGTCATC AGACGGCATC AATTTCGCGG
151 ACGAGCAATT GCCGTTGCTG ATGGAACAAT TGTCGCGCAG CGGTAAAGGCG
201 TTATTGGTCG ATCGGAACGG TCTGTATCTT GCCAACGCCA ATTTCCATCA
251 TGAGGCGGCG GAAGAGTTGG GTTTGTTGGC GGCAAGATC GCACAGATGG
301 AAAAGAAATA CCGGCTGCTG ATTAAGAACA AC..

```

This corresponds to the amino acid sequence <SEQ ID 610; ORF143>:

```

1 MRTKSAVRS CTWADTADID TALNLLYRLQ KLEFLYGDEN GHSDGINLXD
51 EQPLPLMEQL SGSGKALLVD RNLGLYLANAN FHHEAAEELG LLAAEVAQME
101 KKYRLLIKNN ..

```

Further work revealed the complete nucleotide sequence <SEQ ID 611>:

```

1 ATGGAATCAA CACTTTCAC TACAAGCAAT TTATATCCCC GCCTGACTCC
51 TCCCGGTGCA TTTTATGCGG TATCCACGGA TGCCCCAGT GCCGTAATAA
101 CTCTGTGCGA CAGCCTGTGT AAGCAGATG CGGACGAAAT GGTCAGCAGT
151 GAGAAGCTGC TTACTTGGGC GGACACGCC GACATCGATA CCGCTTTGAA
201 CCTGTGTGAC CGTTTGCAAA AACTCGAATT CCTCTATGGC GATGAAACAG
251 GTCATTGAGA CGGCATCAAT TTGTCGAGC AGCAATTGCC GTTGCTGATG
301 GAACAAATGT CGGCGAGCGG TAAGCGGTTA TTGTCGATC GGAACGGTCT
351 GTATCTTGCC AAGCGCAATT TCCATCATGA GCGGCGGAA GAGTTGGGGT
401 TGTGCGGCGC AGAAGTCGCA CAGATGGAAA AGAAATACCG GCTGCTGATT
451 AAGAACAACC TGTATATCAA CAATAACGCT TGGGGCGGTT CGCATCTTTC
501 CGGTCAGAGC GAATTGACAT TTTTCCGATT GTATATCGGT TCAACCAAT
551 TTATTTTGGT TATCGGCGGC ATTCGCGATT TGGGCAAGA GGCATTGTTG
601 ACTTTGGTAA GGATTTTATA CCGCGGTAC AGCAACCGCG TGTA

```

This corresponds to the amino acid sequence <SEQ ID 612; ORF143-1>:

```

1 MESTLSLQAN LYPRLTTPGA FYAVSSDAPS AGKTLHSLL KADADEMVSS
51 EKLLTWADTA DITDALNLLY RLQRLFLYLG DENGSDGIN LSDEQLPLIM
101 EQLSGSGKAL LVDRNGLYLA NANFHHEAAE ELGILLAAEVA QMEKKYRLLI
151 KNNLYINNA WGVCDPSQGS ELTFFFLYIG STKFLVIGG IPDLGKEAFV
201 TLVRILYRRY SNRV*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF143 shows 92.4% identity over a 105aa overlap with an ORF (ORF143a) from strain A of *N.meningitidis*:

				10	20	30
orfl43.pep				MRTKSAVRSCTWADTADIDTALNLLYRLQKLEFL		
orfl43a	GAFYAVSSDXFSAGKTLHSLLLKADADEMVSSEKLLTWADTADIDTALNLLYRLQKLEFL					
	20	30	40	50	60	70

		40	50	60	70	80	90
orf143.pep		YGDENGHS	DGINLXDEQLPL	LMELSGSGK	ALLVDRNGL	YLANANFH	HEAAEELGLLAAE
orf143a		YGDENGHS	DGINLXDEQLPL	LMELSGSGK	ALLVDRNGL	YLANANFH	HEAAEELGLLAAE
5		80	90	100	110	120	130
		100	110				
orf143.pep		VAQMEKKY	RLLIKNN				
orf143a		VAQMEKKY	RLLIKNN				
10		140	150	160	170	180	190

The complete length ORF143a nucleotide sequence <SEQ ID 613> is:

	1	ATGGAATCAA	CANTTTCACT	ACAAGCAAA	TTATATCNCC	GCCTGACTCC
15	51	TGCGCGTGCA	TTTTATGCCG	TATCCAGCGA	TGNCOCAGT	GCCGGTAAAA
	101	CTTTGTTGCA	CAGCCTGTTG	AAGCGGATG	CGGACGAAAT	GSTNAGCACT
	151	GAGAAGCTGC	TTACCTGGGC	GGANACCGCC	GACATCGATA	CCGCTTTGAA
	201	CCTGTTGTAC	CGTTTGCAAA	AACCTGAATT	CCTCTATGCC	GATGAARACG
	251	GTCATTGAGA	CGGCATCAAT	TTGTCGAGG	AGCAATTGCC	GTGCGTGATG
20	301	GAACAATTGT	CCGCGAGCGG	TAGGCGTTA	TTGCGTCATC	GGAAACGTCCT
	351	GTATCTTGCC	AAGCCAAAT	TCCATCATGA	GGCGCGCGAA	GAGTTGSGGT
	401	TGTTGGCGCC	AGAAGTCGCA	CAGATGAGAA	AGAAATACCG	GCTGCNNATT
	451	AAGAACAACC	TGTATATCAA	CAATAACGCT	TGGGCGGCTT	GCATCCITTC
	501	CGGTGAGAGC	GAATTGACAT	TTTTCCCAT	GTATATCGGT	TCAACCAAT
25	551	TTATTTTTGGT	TATCGGCGGC	ATTCCCGATT	TGGGCAAGA	GGCATTTTGTT
	601	ACTTTGGTAA	GGATNTTATA	CNCCNGTTA	CAGCAACCGC	GTGTAAAACT
	651	TGGGAGAGAG	GANGGTTAT	GCAGCAATTA	TTGA	

This encodes a protein having amino acid sequence <SEQ ID 614>:

	1	MESTXSLQAN	LYXRLTPAGA	FYAVSSDXPS	AGKTLHSL	L KADADEMVSS
30	51	EKLTLWAXTA	DIDTALNLLY	RLQKLEFLYG	DENGHS	DGIN LSDEQLPLIM
	101	EQLSGSGKAL	LVDNRGLYLA	NANFHHEAAE	ELGLLAAEVA	QMEKKYRLXI
	151	KNNLYINNNA	WGVCDFSGQS	SLTFEFLYIG	STKFILVIGG	IPDLGKEAFV
	201	TLVRXLYLXL	QQPRVKLGRE	XGLCSNY*		

ORF143a and ORF143-1 show 97.1% identity in 207 aa overlap:

35	orf143a.pep	MESTXSLQANLYXRLTPAGAFYAVSSDXPSAGKTLHSLKADADEMVSSSEKLLTWAXTA
	orf143-1	MESTLSLQANLYPRLTAGAFYAVSSDAPSAAGKTLHSLKADADEMVSSSEKLLTWADTA
40	orf143a.pep	DIDTALNLLYRLQKLEFLYG DENGHS DGINLSDEQLPLLMELSGSGKALLVDRNGLYLA
	orf143-1	DIDTALNLLYRLQKLEFLYG DENGHS DGINLSDEQLPLLMELSGSGKALLVDRNGLYLA
	orf143a.pep	NANFHHEAAEELGLLAAEVAQMEKKYRLXIKNNLYINNNAGWCVDPSPGQSELTFFFLYIG
45	orf143-1	NANFHHEAAEELGLLAAEVAQMEKKYRLIKNNLYINNNAGWCVDPSPGQSELTFFFLYIG
	orf143a.pep	STKFILVIGGIPDLGKEAFVTLVRXLY
	orf143-1	STKFILVIGGIPDLGKEAFVTLVRLYI

50 Homology with a predicted ORF from *N.gonorrhoeae*

ORF143 shows 95.5% identity over a 110aa overlap with a predicted ORF (ORF143ng) from *N.gonorrhoeae*:

55	orf143.pep	MRTKWSAVRSCTWADTADIDTALNLLYRLQKLEFLYG DENGHS DGINLXDEQLPLLMEL	60
	orf143ng	MRTKWSAVRSCSRADTADIDTALNLLYRLQKLEFLYG DENGHS DGINLSDEQLPLLMEL	60
	orf143.pep	SGSGKALLVDRNGLYLANANFHHEAAEELGLLAAEVAQMEKKYRLIKNN	110
	orf143ng	SGSGKALLVDRNGLYLANANFHHEAAEELGLLAAEVAQMEKKYRLIKNNLYINNNAGV	120

An ORF143ng nucleotide sequence <SEQ ID 615> was predicted to encode a protein having amino acid sequence <SEQ ID 616>:

5

1	WRTKTSWASV	CSRADTADID	TALNLYLVRQ	KLEFLYGDEN	GHSNDGNSLSD
51	EQLPLMEQLG	SGSGKALLVD	RGNLYLANAN	FHHESAEELG	LAAEVAQMEK
101	KKYRLLRINR	LYINNNAWQV	RPKSQSGELT	FFPLYIGSTK	FTLVIAIGPD
151	LSKGQCYFPG	KDPTPLPQOW	CVDKSGTGIM	ROLLISTLED	LNNTSTPDIA
201	SAVMTSDGFL	LSMPLPSHLN	SDRVAGTASL	LLAGSRSVQ	ELACGELEGV
251	MIKSGSGYTL	LSQAQDAVL	VLVAKETRIS	GLIILDAKRA	ARHAEAE*

Further work revealed the following gonococcal DNA sequence <SEO ID 617>:

10	1	ATGCGATCAAA	CACITTTTCAT	ACAGCGGGAAT	TATATCCCT	GCTGACTCG
	51	TGCGCGGTGGA	TATTTCCGCG	TATTCACGGA	TGCGCCGAT	GCGGTAAAA
	101	CTTTTGTTCG	CAGCCTGTGT	AAAGCGGTGT	CGGACGAAT	GGTGACGGT
	151	GAGAAGCTGC	TGCGCGCGGA	CACCGCGGAC	CTATGACCG	TTTGAACTC
15	201	GTTGTACGCT	TGCAAAAAAT	TGGAATTCCT	CATTGCGGAT	GAAGACGGT
	251	ATTCAGACGC	GATCAATTTG	TGGAGCAGTC	GTATGCCGTT	GCTGATCGTC
	301	CAATTGTCCG	GCGACGGTAA	GGCATTTATG	GTGCGACGGA	ACGGTCTGTA
	351	CTTGTCCAAAC	GCAATTTTCC	ATCATGATG	GCGGGAAGAG	TGTGGGTTGT
	401	TGCGCGCGAG	AGTGGCACAG	TGGAAGAAAG	TATACCGGCT	GCTGATTCAG
	451	AACACCAACT	ATATCAACAA	TAGCGCTTGG	GCGGCTTTGG	ATCTCTGCGG
	501	TGCAAGATTC	CGGACATTC	TGCAATATTC	TGCAATATTC	ACCAAAATTT
20	551	TGTGTGTAT	CGCGGCATTT	CCGATTTTGA	GCAAAAGGAT	ATTGTTTACT
	601	TGGTAAAGTA	TTTTTAATCC	CCGTTACAGC	ACCGCGGTGT	AA

This corresponds to the amino acid sequence <SEO ID 618: ORF143ng-1>:

25

1	MESTLSLQAN	LYPELTPAGA	FYAVSSDAPS	AGKTLRLSL	KADADEVVSS
51	EKLLAADATD	IDTALLNLYR	LQKLEFLYGD	ENGHSDGINL	SDEQLPLLME
101	QLSGSGKALL	VDRNGLYLAN	ENHFEHSAAE	LGLLAAEVAQ	MEKKYRLLRL
151	NNLYINNNAA	GVCPDPSQSE	LTFFPLYTGS	TKFILVIAGI	PDLSEKFAVT
201	LURLYREYS	NRV*			

ORF143ng-1 and ORF143-1 show 95.8% identity in 214 aa overlap:

30	orf143ng-1.pep	MESTLSLQANLYPCLTPAGAFYAVSSDAPSGAKTLLRSLKADADEVSVSEKLLA-ADTA	59
	orf143-1	MESTLSLQANLYPRLTPAGAFYAVSSDAPSGAKTLLRSLKADADEMVSSEKLLIWDATA	60
35	orf143ng-1.pep	DIDTALNLLVYRQKLEFFLYGDENGHSDGINSLSDEQLPLMEQLSGSGKALLVDNRNGLYLA	119
	orf143-1	DIDTALNLLVYRQKLEFFLYGDENGHSDGINSLSDEQLPLMEQLSGSGKALLVDNRNGLYLA	120
40	orf143ng-1.pep	NANFHHEAAEELGLLAARVAQMEKKYRLLRINNLINNNAWGVCDPSGQSELTFFPLYIG	179
	orf143-1	NANFHHEAAEELGLLAARVAQMEKKYRLLRINNLINNNAWGVCDPSGQSELTFFPLYIG	180
	orf143ng-1.pep	STKFLIVIAIGIPDLKSKEAFVTLLVRIIVRYRNSRV	213
	orf143-1	STKFLIVIGGIPDLGKEAFVTLLVRIIVRYRNSRV	214

45 Based on the presence of the putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 74

The following partial DNA sequence was identified in *N.meningitidis* <SEO ID 619>:

50 1 ATGACCTTTT TACAACGTTT GCAAGGTTTG GCAGACAATA AAATCTGTGC
 51 GTTTGCAATGG TPCGTCTGCC GCCGCTTTGA TGAAGAACGC GTACCCGAGr
 101 CGGCGGCAAG CATGACGTTT ACGACGCTGC TGGCACTCGT CCCCGTGCTr

5
 151 ACCGTGATGG TGGCGGTGCG TTCGATTTTC CCCGTGTTGG ACCGCTGGTC
 201 GGATTGCTTC GTCTCCTTCG TCAACCAAC CATTGTGCGG CA.GGGCGGG
 251 ACATGGTGTT CGACTATATC AATGCGTTCC GCGAGCAGGC GAACCGGCTG
 301 ACGCAATACG GCAGCGTGAT GCTGGTTCGT ACCTCGCTGA TGCTGATTCG
 351 GACGATAGAC AATACGTTCA ACCGATCTCG GCGGGTCAA WTYCCAGGCT
 401 CCGTGGATG..

This corresponds to the amino acid sequence <SEQ ID 620; ORF144>:

10
 1 MTFLQRLQGL ADNKICAFW FVVRFDDEER VPQAAASMTF TTLALVLPVL
 51 TVMVAVASIF PVFDRWSDSF VSEVNTIVP XGADMFVDYI NAFREQANRL
 101 TAIGSVMLVV TSLMLIRTID NTFNRWRVW XQRPMW...

Further work revealed the complete nucleotide sequence <SEQ ID 621>:

15
 1 ATGACCTTTT TACAACGTTT GCAAGGTTTG GCAGACAATA AAATCTGTGC
 51 GTTTGCGATGG TTGTCGTCCT GCGCGTTTGA TGAAGAACGC GTACCCGAGG
 101 CGGCGGCAAG CATGACGTTT ACGACGCTGC TGGCACTCGT CCCGCTGCTG
 151 ACCGTGATGG TGGCGGTGCG TTTCGATTTTC CCGCTGTTCC ACCGCTGGTC
 201 GGATTGCTTC GTCTCCTTCG TCAACCAAC CATTGTGCGG CAGGCGCGGG
 251 ACATGGTGTT CGACTATATC AATGCGTTCC GCGAGCAGGC GAAACCGGCTG
 301 ACGGCAATCG GCAGCGTGAT GCTGGTTCGT ACCTCGCTGA TGCTGATTCG
 351 GACGATAGAC AATACGTTCA ACGCATCTCG GCGGGTCAAT TCCCAGCGTC
 20
 401 CCGGGATGAT GCGATTCTTC GTCTATTGGG CTTTACTGAC GTTCGGGCGG
 451 CTGCTCTTGG GCGTGGGCAT TTCTGTTATG TCCGGCTGCG TACAGGATGC
 501 CCGCGTTGCC TCAGGTGCGC CGCAGTGCTG GCGCGGCTGG CGAACGGCGG
 551 CGACGCTGAC CTTCATGAGC CTTTGTGCTG GGGGCGCTGA CCGCTCTGCTG
 601 CCAACCGGCT TCGTCCCGC GGGCGAGCGC TTTGTGGGG CTTCGGCAAC
 25
 651 AGCGTTTGTG CTGGAAACCG CGCGCTCCCT CTTCATCTGG TATATGGCA
 701 ATTTTCAGCG CTACCGCTCG ATTTACGGCG CTTTCCGCG CGTGCGGTTT
 751 TTTCTGTTGT GGCTGAACCT GTTGTGGACG CTGCTCTGG GGGCGCGGCT
 801 GCTGACTTCT TCACCTCTCT ACTGGCAGG AGAAGCGTTC CGCAGGGGCT
 851 TCGACTCGCG CGGACGGTTT GACGACGCTG TGAATACTCT GGTGCTTCTG
 901 GATCGGCGCG AAAAAGAAGG CAAAGCCTTG CCGTTCAGG AGTTTCAGACG
 951 GCATATCAAT ATGGGCTAAG ACGAGTTGGG CGAGCTTTTG GAAAAGCTGG
 1001 CGCGGCAAGG CTACATCTAT TCCGGCAGAC AGGCTTTGGT GTTGAACACG
 1051 GGGCGGATTT CGATTGAGTT GAACGAACCT TCAAGCTCT TCGTTTACCG
 1101 TCCCTTGCCT CTGGAAAGGG ATCATGTGAA CCAAGCTGTC GATGCGGTAA
 35
 1151 TGACACCGTG TTTGCAACT TTGAACATGA CGTGGCGAGA GTTTGACGCT
 1201 CAGGCGAAAA AACGCGACTA G

This corresponds to the amino acid sequence <SEQ ID 622; ORF144-1>:

40
 1 MTFLQRLQGL ADNKICAFW FVVRFDDEER VPQAAASMTF TTLALVLPVL
 51 TVMVAVASIF PVFDRWSDSF VSEVNTIVP XGADMFVDYI NAFREQANRL
 101 TAIGSVMLVV TSLMLIRTID NTFNRWRVW XQRPMWQFL VYWALLTFGP
 151 LSLGVGISFM VGSVQDAALA SGAPQWSGAL RTAATLTFMT LLLWGLYRFV
 201 PNRFVVARQA FVGALATAFC LETARSLFW YMGNFDGYRS IYGAFVAVPV
 251 FLLNLNLWLT LVLGGAVLTS SLSYWGQEF RRGFDSRGFR DDVLKILLLL
 45
 301 DAAQKEGKAL PVQEFRRHIN MGYDELGELL EKRLARHGYY SGRQGWVLKT
 351 GADSIENLNL FKLFIYRELP VERDHVNGAV DAVMTPLCLQ INMTLAEFDA
 401 QAKRQ*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF144 shows 96.3% identity over a 136aa overlap with an ORF (ORF144a) from strain A of *N.*

meningitidis:

50
 55
 orf144.pcp MTFLQRLQGLADNKICAFWVVRFDDEERVPOXAAASMTFTTLALVLPVLTVMVAVASIF
 orf144a MTFLQRLQGLADNKICAFWVVRFDDEERVPOXAAASMTFTTLALVLPVLTVMVAVASIF
 10 20 30 40 50 60
 70 80 90 100 110 120


```

    orf144.pep    PVFDRWSDSFVSVFNQTI VXPQADMVFDYINAFREQANRLTAIGSVMLVVTSLMLIRTID
    orf144a      EVFDRWSDSFVSVFNQTI VXPQADMVFDYINAFREQANRLTAIGSVMLVVTSMXLMIRTID
                70      80      90      100      110      120

5
    orf144.pep    130
    orf144a      NTFNRIWRVNSQRPWMMQFLVYWALLTFGPLSLGVGISFXVSVQDAAALASGAPQWSGAL
                130      140      150      160      170      180

```

The complete length ORF144a nucleotide sequence <SEQ ID 623> is:

```

1   ATGACCTTTT TACAACGTTT GCAAGGTTTG GCAGACAATA AAATCTGTGC
5   GTTTGCATGG TTCGTCGTCC GCGGCTTTGA TGAAGAACGC GTACCCGACG
10  CCGCGCGCAAG CATGACGTTT ACGACACTGC TGGCACTCGT CCCCCTGCTG
15  ACCGTGATGG TGGCGGTGCG TCGGATTTTC CCGGTGTCG ACCGNTGGTC
20  GGATTCGCTG GTCTCCTTCG TCAACCAAC CATTTGTCCG CAGGGCGGGC
25  ACATGGTNTT GCATATATC AATGCGTTCC GCGAGCAGGC GAACCGGCTG
30  ACGGCAATCG GCAGCGTGAT GCTGGTGCTT ACCTCGCNGA TGCTGATTGG
35  GACGATAGAC AATACGTACA ACCGCACTCG CCGGCTCAAT TCCGACGCTC
40  CTTGGAGATG GACTCTTCTC GTCTATTGGG CTTTACTGAC GTTCCGGGCG
45  CTGTCTTTGG GCGTGGGCACT TTCCTTTATN GTCCGCTCGG TACAGAGTGC
50  CCGCGCTGCC TCAGGTGCGC CGCAGTGGTC GCGCGGCTTG GCAACGGCGG
55  CGACGCTGAN CTTTCATGAG CTTTGTGCTG GGGGCGTGTA CCGCTNCGTG
60  CCAAAACCGCT TCGTTCCTCG GCGGCANGCG TTTGTCGGGG CTTTGGCAAC
65  AGCGTTCTGT CTGGAACACG CGCGTTCCTT CTTTACTTGG TATATGGGCA
70  ATTTCGACGG CTACCGCTCG ATTTACGGNG CGTTTCGCGC CGTGCGGTTT
75  TTTCTGTGTT GGCTGAACCT GTTGTGGACG CTGGCTCTGG CCGCGCGCGT
80  GCTGACTTCT TCACTCTCCT ACTGGCAGGG AGAAGCGTTC CGCAGGNGCT
85  TCGACTCGCG CGGACGCGTT GACGACGTGT TGAAAATCCT GCTGCTCTG
90  GATCGCGCGC AAAAAGAAGG CNAAGCCTTG CCTGTTCAAG AGTTCAAGCG
95  GCATATCAAT ATGGGCTACG ACGAGTTGGG CGAGCTTTTG GAAAAGCTGG
100 CCGCGCACGG CTACATCTAT TCCGCGCAGC AGGGTTGGGT TCCGAAAACG
105 GCGCGCGAGT CGATTGAGTT GAACGACATC TTCAGCTCTC TCGTTTACCG
110 TCGCTTGCTT GTGGAAGGGS ATCATCTGAA CCAAGCTGTC GATGCGGTAA
115 TGATGCGCGT TTTGCAGACT TTGAACATGA CGCTGGCAGA GTTTAGCGCT
120 CAGCGCAAAA AACAGCAGCA ATCTTGA

```

This encodes a protein having amino acid sequence <SEQ ID 624>:

```

1   MTFLLRLQGL ADNKICAFAW FVVRRFDEER VPQAAASMTF TLLALVPVL
5   TVMVAVASIF PVFDRWSDSF VSVFNQTI V P QADMVFDYI NAFREQANRL
10  TAIGSVMLVV TSXLMIRTID NTFNRIWRVN SQRPWMMQFL VYWALLTFGP
15  LSLGVGISFX VGSVQDAAALA SGAPQWSGAL RTAATLXFMT LLLWGLYRKV
20  PNRFPVPRXA FVGALATAFC LETARSLFTW YMGNFDGYS IYGAFAPVFP
25  FLLMLNLLWT LVLGGAVALTS SLSYWQGEAF RRXFPDSRGF DDVLKILLL
30  DAAQREGKAL EVQEPFRHIN MGYDELGELL EKLARHGYIY SGRQGWVLKT
35  GADSIELNEL FKLFRVIRFLP VERDHNQAV DAVMMFCIQT LNMTLAEFDA
40  QAKKQQQS*

```

ORF144a and ORF144-1 show 97.8% identity in 406 aa overlap:

```

    orf144a.pep    MTFLLRLQGLADNKCICAFANFVVRRFDEERVQPAAASMTFTTLLALVPVLVIMVAVASIF
    orf144-1      MTFLLRLQGLADNKCICAFANFVVRRFDEERVQPAAASMTFTTLLALVPVLVIMVAVASIF
    orf144a.pep    PVFDRWSDSFVSVFNQTI VXPQADMVFDYINAFREQANRLTAIGSVMLVVTSLMLIRTID
    orf144-1      PVFDRWSDSFVSVFNQTI VXPQADMVFDYINAFREQANRLTAIGSVMLVVTSLMLIRTID
    orf144a.pep    NTFNRIWRVNSQRPWMMQFLVYWALLTFGPLSLGVGISFXVSVQDAAALASGAPQWSGAL
    orf144-1      NTFNRIWRVNSQRPWMMQFLVYWALLTFGPLSLGVGISFXVSVQDAAALASGAPQWSGAL
    orf144a.pep    RTAATLXFMTLLLWGLYRKVPNRFVPRKAFVGAALATAPCLETARSLFTWYMGNFDGYS
    orf144-1      RTAATLTFMTLLLWGLYRFPNRFVPRKAFVGAALATAPCLETARSLFTWYMGNFDGYS

```

	orf144a.pep	IYGAFAAVPFFLLWNLNLTVLGGAVLTSSLSYWGGEAFRRXFDSRGRFDDVLIKILL	
	orf144-1	IYGAFAAVPFFLLWNLNLTVLGGAVLTSSLSYWGGEAFRRXFDSRGRFDDVLIKILL	
5	orf144a.pep	DAAQKEGKALPVQEFRRHINMGYDELGELLEKLARHGYIYSGRQGVWLTGADSIENEL	
	orf144-1	DAAQKEGKALPVQEFRRHINMGYDELGELLEKLARHGYIYSGRQGVWLTGADSIENEL	
10	orf144a.pep	FKLFVYRPLPVERDHVNQAVDAVMMPCQLTNMTLAEFDAQAKQQQS	408
	orf144-1	FKLFVYRPLPVERDHVNQAVDAVMMPCQLTNMTLAEFDAQAKRQ	406

Homology with a predicted ORF from *N.gonorrhoeae*

ORF144 shows 91.2% identity over a 136aa overlap with a predicted ORF (ORF144ng) from *N.gonorrhoeae*:

	orf144.pep	MTFLQRLQGLADNKICAFAWFVVRFDSEERVPOXAASMTFTTLLALVPVLTMVAVASIF	60
	orf144ng	MTFLQCWQGSADNKICAFAWFVIRFSEERVPOXAASMTFTTLLALVPVLTVMVAVASIF	60
20	orf144a.pep	PVFDRWSDSFVSFVNQITVXPQADMVFDYINAFREQANRLTAIGSVMLVVTSLMLIRTID	120
	orf144ng	PVFDRWSDSFVSFVNQITVXPQADMVFDYIDAFRQANRLTAIGSVMLVVTSLMLIRTID	120
25	orf144.pep	NTFNRIWRVXXQRPWM	136
	orf144ng	NAFNRIWRVNTQRFWMQFLVYWALLTFGLPSLGVGISFVMVGSVDLSVSSAQGWADAL	180

The complete length ORF144ng nucleotide sequence <SEQ ID 625> is predicted to encode a protein having amino acid sequence <SEQ ID 626>:

30	1	MTFLQCWQGS	ADNKICAFAW	FVIRFSEER	VPQAAASMTF	TTLLALVPVL
	51	TVMVAVASIF	PVFDRWSDSF	VSFVNQITVP	QGADMVFDYI	DAFRDQANRL
	101	TAIGSVMLVV	TSLMLIRTID	NAFNRIWRVN	QRFWMQFL	VYWALLTGEF
	151	LSLQGVISFM	VGSVDVSLVS	SGAQWADAL	KTARLAEFTW	LLWGLYRFV
	201	PNRFVFAQQA	FVGALITAFG	LETARELFTW	VMGNFDFVRS	IYGAFAAVPE
35	251	FLWLNLNLT	VLVGGAVLTS	SLSYWGGEAF	RREGFDSRGP	DVVLKILLLL
	301	DAAQKEGRTL	SVQEFRRHIN	MGYDELGELL	EKLARVGYIY	SGRQGVWLTG
	351	GADSIELSEL	FKLFVYRPLP	VERDHVNQAV	DAVMTFCLQT	INMTLAEFDA
	401	QAKKQQQS*				

Further work revealed the following gonococcal DNA sequence <SEQ ID 627>:

40	1	ATGACCTTTT	TACAAAGCTG	GCAAGGTTTG	GCGGACAATA	AAATCTGTGC
	51	ATTTCATGCG	TTCGTCATCC	GCCGTTTCAG	TGAAGAGCGG	GTACCCGAGG
	101	CAGCGCGGAG	CATGACGTTT	ACGACACTGC	TGGCACTCGT	CCCCGTACTG
	151	ACCGTAATGG	TGCGGCTCGC	TTCGATTTCG	CCCGTGTTCC	ACCGCTGGTC
	201	GGATTCGTTT	GTCTCCTTCG	TCAACCAAAAC	CATTGTGCGC	CAGGCGCGCG
45	251	ATATGGCTTC	CGACATATTC	CGACATATTC	CCGATCAGCG	AAACCGGCTG
	301	ACCGCCATCG	CGAGCGTGAT	CGTGCCTCTA	ACCTCGCTTAC	TGCTGATTCG
	351	GACGATAGAC	AATGCGTTCA	ACCGCATCTG	CGCGGCTTAC	ACCGAACGCC
	401	CCTGGATGAT	CGAGTTCCCT	GTTTATTGGG	CGTTCGTGAC	TTTCGGGCGT
	451	TGTGCTTTGG	GTGTGGGCGT	TTCCTTTATG	GTCCGGTCGG	TTCAGACTCG
	501	CGTACTCTCC	TCCGGAGCGG	AACAATGGCG	GGACCGCTTG	AAGACGCGCG
50	551	CAAGGCTGGC	TTTCATGACG	CTTTGCTGTG	GGGGCGGTGA	CCGCTTCGTG
	601	CCCAACCGCT	TCGTGCCCGC	CGCGCAGGCG	TTTGTGGAGG	CTTTGATTAC
	651	GGCATTCTCG	CTGGAGACGG	CACGTTTCCT	GTTCACCTGG	TATATGGGCA
	701	ATTTTCGACG	CTACCGCTCG	ATTATCGGCG	CATTGTCCCG	CGTGCCGTTT
55	751	TTCTCGTCTG	GGTTAAACCT	CGTGTGGAGC	CTGGTCTTGG	CGCGGCGCGT
	801	CGTGACTCTG	TGCGTGTCTT	ATTGGCAGGG	CGAGGCGCTT	CGCAGGGGAT
	851	TGACTCTCGG	CGGACGGTTT	GACGACGTGT	TGAAAACTCT	CGTGCTTCTG
	901	CATGGCGCGC	AAAGAGAGG	CCGACCCCTG	TCGGTTCAGG	AGTTACAGAG
	951	CGATATCAAT	ATGGGTTTACG	ATGATATGAG	CGAGCTTTTG	GAAAGGCTGG
60	1001	CGCGGTACGG	CTATATCTAT	TCCGCGACAG	AGGGCTGGGT	TTTGAAACAG
	1051	GGGGGGGATT	CGATTGAGTT	GACCGAAGCT	TTCAAGCTCT	TCGTGTACGG

This encodes a variant of ORF144ng, having the amino acid sequence <SEQ ID 628; ORF144ng-

1	MTFLFRQWQGL	DNHKKICAFAP	VYFRIRRSSE	VOQAASMTST	TLTLLALVPL
51	TUMVAVASEF	PVFERDWSDF	VSVFNTQTVF	QGDQMVFDYI	DAFRKQDANKRL
101	TAIGSVMLVF	TSMLIRTID	NAFNRNWRVN	TQORWMMXG	VYVALLTFRFG
151	LSLGVSVLSE	VGSVGSVLS	GQAQWADAL	KTAARLAFMT	LLLLNGLVXFP
201	PNKRFVAPQA	FVLGATATC	LETARLFTF	YMGNFQDGYS	IVGAFAAVPEF
251	FLWLNLMLLT	LVLGAVLTS	SLSYWGCAF	RGMSRGSGE	TDGLALLLL
301	DAQREGRTL	SQVEARRHM	VERDHGELL	EXARVQVGL	SGRWVWVGL
351	QKQKQKQK*	FKLFWYELP	WQNDHNGQV	DAVMTPECLT	LNNLTLSFFDA

orf144ng-1.pep MTFLQRQGLQADNKKICAFAWFVIRFSEERVQQAASMTFTTTLLALVPLVITVMVAVASIF
orf144-1 MTFLQRQGLQADNKKICAFAWFVRRFDEERVQQAASMTFTTTLLALVPLVITVMVAVASIF
orf144ng-1.pep PVFDNRWSDSFVSVFNQYIIPVPGADMVFYDIADFRRQANRLTAIGSVMLVVTSLMLIRTID
orf144-1 PVFDNRWSDSFVSVFNQYIIPVPGADMVFYDINAFREQANRLTAIGSVMLVVTSLMLIRTID
orf144ng-1.pep NAFNRIRWNTQRPWMMQFLVYWYALLTFGLPLSLGVGISFMVGSVQDSVLSSGAQQWADAL
orf144-1 NTFNRIRWVNSQRPMWMMQFLVYWYALLTFGLPLSLGVGISFMVGSVQDAALASGAPQWGGAL
orf144ng-1.pep KTAARLAFMTLLLGWLYRFVPNRFVPAQAFVFGALITAFCLETARLFTWYMGNFQGYRS
orf144-1 RTAATLTFTMTLLLGWLYRFVPNRFVPAQAFVFGALITAFCLETARSLFTWYMGNFQGYRS
orf144ng-1.pep IYGAFAVVPFFLLWLNLWLTVLGGAVLTTSSLSYWQGEAFRRGFDSSRGFPDVLKILL
orf144-1 IYGAFAVVPFFLLWLNLWLTVLGGAVLTTSSLSYWQGEAFRRGFDSSRGFPDVLKILL
orf144ng-1.pep DAAQKEGRTLSVQGEFRHHNMCGYDELGELEKLARGHYIYSGRQGWILKTGADSIENEL
orf144-1 DAAQKEGKGLFVQGEFRHHNMCGYDELGELEKLARGHYIYSGRQGWILKTGADSIENEL
orf144ng-1.pep FKFLEVYRPLPVERDHVNQAVDAVMTFCIQTLLNMTLAEFDAQAKKQQQS
orf144-1 FKFLEVYRPLPVERDHVNQAVDAVMTFCIQTLLNMTLAEFDAQAKKQ

On this basis of this analysis, including the identification of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 629>:

```

1  ..AGACACGCC  GCCGCATCCG  CATCGACACC  GCCATCAACC  CCGAACTGGA
51  AGCCCTCGCC  GAACACCTCC  ACTACCAATG  GCAGGGCTTC  CTCTGGCTCA
101  GCACCGATAT  GCGTCAGGAA  ATTTCCGCC  TCGTCATCCT  GCTGCAACGC
151  ACCCGCGGAA  AATGGCTGGA  TGCCCAACGA  CGCCAACACC  TGGCCCAAAG
201  CCTGCTCGAA  ACACGGTGA  ACGGGCTGA

```

This corresponds to the amino acid sequence <SEQ ID 630; ORF146>:

```

1  ..RHARRIRIDT AINPELEALA EHLHYQWGF LWLSTDMRQE ISALVILLQR
51 TRRKWLDAHE RQHLRQSLLE TREHG*

```

Further work revealed the complete nucleotide sequence <SEQ ID 631>:

```

5      1  ATGAACACCT  CGCAACGCAA  CGGCCTCGTC  AGCCGCTGGC  TCAACTCCTA
51     CGAACGCTAC  CGCTACCGCC  GCCTCATCCA  CGCCGTCGGG  CTCGGCGGGG
101    CGCTCCTGTT  CGCCACCGCC  TCCGCGCGGC  TGCTCCACCT  CCAACACGGC
151    GAGTGGATAG  GGATGACCGT  CTTGCTCGTC  CTCGGCATGC  TCCAGTTTCA
201    AGGGGGGATT  TACTCCAAGG  CGGTGGAACG  TATGCTCGGC  ACGGTTCATG
10     251  GGCTGGGGCGC  GGTTTGGGCG  GTTTTATGGC  TGAACACGCA  TTATTTCCAC
251    GCGCAACTCC  TCCTTCTACT  CACCGTCGGC  ACGCGAAGCG  CACTGCGCGG
351    CTGGGGCGCG  CTCGGCAAAA  ACGGCTACGT  CCTATPCTGT  GCAGGGCTGA
401    CGATGTGTAT  GCTCATCGGC  GACAACGGCA  GCGAATGCTG  GCACACGGGA
451    CTCATGCGCG  CCATGAACGT  CCTCATCGGC  CGCGGCATCG  CCATGCGCGC
15     501    CGCCAAACTG  CTCGCGCTGA  AATCCAACT  GATGTGGCGT  TTCTATGCTT
551    CGGACAACCT  GGCGGACTGC  AGCAAAATGA  TTGCGGAAAT  CAGCAACGGC
601    AGGCGCATGA  CCGCGGAACG  CCTCGAGGAG  AACATGCGCA  AAATGCGCCA
651    AATCAACGCA  CGCATGGTCA  AAGCGCGCAG  CCATCTCGCC  GCCACATCGG
701    GCGAAGCGCG  CATCAGCCCC  GCCATGATGG  AAGCCATGCA  GCACGCCCCA
751    CGTAAATTCG  TCAACACCAC  CGAGCTGCTC  CTGACCACCG  CGGCCAAGCT
801    GCAATCTCCC  AAATCAACG  CGAGCGAAT  CCGGCTGCTT  GACCGCACT
851    TCACACTGCT  CAAAACCGAC  CTGCACACAA  CGGTGCGCCT  TATCAACGGC
901    AGACACGCCC  GCCGCTACCG  CATCGACACC  GCCATCAACC  CCGAACCTGA
951    AGCCCTCGCC  GAACACCTCC  ACTACCAATG  CAGAGGCTTC  CTCGGCTCTG
25     1001   GCACCAATAT  GCGTACGAA  ATTTCGCGCC  TCGTCACTCT  CTGTCACGCG
1051   ACCGCGCGCA  AATGCTGCGG  TGCCACACAA  CGGCCAACAC  TGCGCCAAAG
1101   CCTGCTTGA  ACACGGGANC  ACGGCTGA

```

This corresponds to the amino acid sequence <SEQ ID 632; ORF146-1>:

```

30      1  MNTSQRNRLV  SRWLNSYERY  RYRRLIHAVR  LGAVLFATA  SARLLHLQHG
51     EWIGMTVFVV  LGLMQFOGAI  YSKAVERMLG  TVIGLGAGLG  VLWLNQHYHF
101    GNLLFYLTVG  TASALAGWAA  VGNKNYVPM  AGLTMCMLIG  DNGSEWLDSG
151    LMRAMNVLIG  AATAIAAACL  LPLKSTLMWR  PMLAANLADC  SKMIAETISNG
201    RRMTRERLEE  NMAKMRQINA  RMVKSRLSLA  ATSGESRISP  AMMEAMQHAI
251    RKIVNTTELL  LTAAAKLQSP  KLNQSEIRLL  DRHFTLLQTD  LQQTVALING
35     301    RHARRIRIDT  AINPELEALA  EHLHYQWGF  LWLSTNMRQE  ISALVILLQR
351    TRRKWLDAHE  RQHLRQSLLE  TREHG*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF146 shows 98.6% identity over a 74aa overlap with an ORF (ORF146a) from strain A of *N.*

40 *meningitidis*:

```

                                     10      20      30
orf146.pep                                RHARRIRIDTAINPELEALAEHLHYQWGF
                                     |||
45 orf146a      KLNGSEIRLLDRHFTLLQTDLQQTVALINGRHARRIRIDTAINPELEALAEHLHYQWGF
                280      290      300      310      320      330
                                     40      50      60      70
orf146.pep      LWLSTDMRQEISALVILLQRTTRKWLDAHERQHLRQSLLETREHG
                |||
50 orf146a      LWLSTNMRQEISALVILLQRTTRKWLDAHERQHLRQSLLETRHSX
                340      350      360      370

```

The complete length ORF146a nucleotide sequence <SEQ ID 633> is:

```

55      1  ATGAACACCT  CGCAACGCAA  CGGCCTCGTC  AGCCGCTGGC  TCAACTCCTA
51     CGAACGCTAC  CGCTACCGCC  GCCTCATCCA  CGCCGTCGGG  CTCGGCGGGG
101    CGCTCCTGTT  CGCCACCGCC  TCCGCGCGGC  TGCTCCACCT  CCAACACGGC
151    GAGTGGATAG  GGATGACCGT  CTTGCTCGTC  CTCGGCATGC  TCCAGTTTCA

```

201 AGGGGGCGATT TACTCCAAGG CGGTGGAACG TATGCTCGGC ACGGTCACTG
 251 GGCTGGGCGC GGGTTTGGGC GTTTTATGGC TGAACCAGCA TTATTTCCAC
 301 GGCACCACTCC TCTTCTACCT CACCGTCGCG ACGGCAAGCG CACTGGCGGG
 351 CTGGGGGGCGG GTCGGCAAAA ACGGCTACGT CCTATGCTG GCGGGGCTGA
 401 CGATGTGCAT GCTCATCGGC GACAAACGCA CGCAATGCT GCGACGCGCG
 451 CTGATCGCGG CGATGAACCT CCTCATCGCG GCGGCGATCG CCATCGCGCG
 501 CCGCAAACTG CTGCGCGTGA AATCCACACT GATCTGGCT TTCATGCTTG
 551 CCGACAACCT GACCGACTGC AGCAAAATGA TTGCGGAAAT CAGCAACGCG
 601 AGCGCGCATGA CCGCGGAACG CCTCGAAGAG AACATGCGCA AAATGCGCCA
 651 AATCAACGCA CGCATGGTCA AAGCGCGAG CCACCTCGCC GCCACATCGG
 701 GCGAAGCCG CATCAGCCCC GCGATGATGG AAGCATGCA GCACGCCAC
 751 CGTAAATTG TCAACACCAC CGAGCTGCTC CTGACCACCG CCGCCAAGCT
 801 GCAATCTCCC AAATCAACG CGACGGAAT CGGCTGCTT GACGCCACT
 851 TCACACTGCT CCAAACCGAC CTGCAACAAA CGTTCGCCCT TATCAACGCG
 901 AGACACGCC CCGCATCCG CATCGACACC GCGCATCAAC CCGAATCGGA
 951 AGCCCTCGCC GAACACTCC ACTACCAATG GCAGGCTTC CTCTGSGCTCA
 1001 GCACCAATAT GCGTCAGGAA ATTTCCGCC TCGTATCCT GGTGCAACGC
 1051 ACCGCGCGCA AATGGCTGGA TGCCACAGAA CGCCAACACC TGCGCCAAG
 1101 CCGTCTTGA ACACGGGAAC ACAGTTGA

20 This encodes a protein having amino acid sequence <SEQ ID 634>:

1 MNTSQNRNLV SRWINSYERY RYRRLIHAVR LGAGVLFATA SARLLHLQHG
 51 EWIGMTVEVV LGMLQFGAI YSKAVERMLG TVIGLGAGLV VLWLNQHYFH
 101 GNLFYLVTVG TASALAGWAA VKNNGYVPM LAGTMCMLIG DNGSEWFDG
 151 IMRAMNVILG AAIATAAKL LPLKSTIMWR FMLADNLTD SKMIAEISNG
 201 RMRTRERLEE NMAKMRQINA RMVKSRSLHA TSGESRISP AMMEAMQHAH
 251 RKIIVNTTELL LTTAAKLQSP KLNGSEIRLL DRHFTLLQTD LQQTVALING
 301 RHARRIRIDT AINFELEALA EHLHYQWQGF LWLSTNMRQE ISALVILLQR
 351 TRRKWLDAHE RQHLRQSLL E TREHSX*

ORF146a and ORF146-1 show 99.5% identity in 374 aa overlap:

30 orf146a.pep MNTSQNRNLVSRWINSYERYRYRRLIHAVRLGGAVLFATASARLLHLQHGSEWIGMTVEVV
 orf146-1 MNTSQNRNLVSRWINSYERYRYRRLIHAVRLGGAVLFATASARLLHLQHGSEWIGMTVEVV
 35 orf146a.pep LGMLQFGCAIYSKAVERMLGTVIGLGAGLVVLWLNQHYFHGNLLFYLVTVGTASALAGWAA
 orf146-1 LGMLQFGCAIYSKAVERMLGTVIGLGAGLVVLWLNQHYFHGNLLFYLVTVGTASALAGWAA
 40 orf146a.pep VKNNGYVPM LAGTMCMLIGDNGSEWFDGSLMRAMNVILGAAIAIAAAKLLPLKSTIMWR
 orf146-1 VKNNGYVPM LAGTMCMLIGDNGSEWFDGSLMRAMNVILGAAIAIAAAKLLPLKSTIMWR
 45 orf146a.pep FMLADNLTDCKSMIAEISNGRMRTRERLEENMAKMRQINARMVKSRSLHAATSGESRISP
 orf146-1 FMLADNLTDCKSMIAEISNGRMRTRERLEENMAKMRQINARMVKSRSLHAATSGESRISP
 50 orf146a.pep AMMEAMQHAHRKIIVNTTELLTTAAKLQSPKLNGSEIRLLDRHFTLLQTDLQQTVALING
 orf146-1 AMMEAMQHAHRKIIVNTTELLTTAAKLQSPKLNGSEIRLLDRHFTLLQTDLQQTVALING
 55 orf146a.pep RHARRIRIDTAINFELEALAEHLHYQWQGFVLWSTNMRQEISALVILLQTRRRKWLDAHE
 orf146-1 RHARRIRIDTAINFELEALAEHLHYQWQGFVLWSTNMRQEISALVILLQTRRRKWLDAHE
 60 orf146a.pep RQHLRQSLL E TREHSX
 orf146-1 RQHLRQSLL E TREHGX

Homology with a predicted ORF from *N.gonorrhoeae*

ORF146 shows 97.3% identity over a 75aa overlap with a predicted ORF (ORF146ng) from

N.gonorrhoeae:

```

      RHARRIRIDTAINPELEALAEHLHYQWQGF 30
orf146.pep  |||
orf146ng    KLNGSEIRLLDRHFTLLQTDLQQTAAALINGRHARRIRIDTAINPELEALAEHLHYQWQGF 364
5          |||
orf146.pep  LWLSTNMRQEISALVILLQRTRRKWLDAHERQHRLRQSLLETREHG 75
      |||
orf146ng    LWLSTNMRQEISALVILLQRTRRKWLDAHERQHRLRQSLLETREHG 409

```

An ORF146ng nucleotide sequence <SEQ ID 635> was predicted to encode a protein having amino acid sequence <SEQ ID 636>:

```

10      1  MSGVRFPSPA  PIPSTDPSPG  SLCFFTFPQ  TASDMNSSQR  KRLSGRWLNS
      51  YERYRHRLLI  HAVRLGGTVL  FATALARLLH  LQHGIEWIGMT  VFVVLGMLQF
      101  QGATYSNAVE  RMLGTVIGLG  AGLGVWLWLNQ  HYFHNLLFY  LTIGTASALA
      151  GWAAVGKNYG  VPMLAGLTMC  MLIGDNGSEW  LDSGLMRAMN  VLIGAAIAIA
      201  AAKLLPLKST  LMWRFMLADN  LADCSKMIAE  ISNRRMTRE  RLEQNMVKMR
      251  QTNARMVKSR  SHLAATSGES  RISPMSMEM  QAHARRIVNT  TELLITTAAK
      301  LQSPKLNGSE  IRLDRHFTL  LQTDLQQTAA  LINGRHARRI  RIDTAINPEL
      351  EALAEHLHYQ  WQGFWLSTN  MRQEISALVI  PLQRTRRKWL  DAHERQHRLQ
      401  SLLETREHG*

```

Further work revealed the following gonococcal DNA sequence <SEQ ID 637>:

```

20      1  ATGAACCTCT  CGCAACGCAA  ACGCCTTTCC  GgccGCTGCG  TCAACTCCTA
      51  CGAACGCTTAc  cGCCAaccGCC  GCCTCATACA  TGCCGTGCGG  CTCGGGggaa
      101  ccgTCCTGTT  GCCCACCGCA  CTCGCCCGgc  TACTCCACCT  CCAacacggc
      151  gAATGGATAG  GGATgaCGGT  CTTCGTCGTC  CTCGSGATGC  TCCAGTTCaA
      201  AGGCgcgatt  Lactccaacg  cggTgaacg  taTgctcggt  accgTcatcg
      251  ggcTgGgcgc  GGGTTGdgcc  gTTTTATGCG  TGAACACGCA  TTAttccac
      301  ctGGGCGCGCG  GTCGCGCAAA  accgctacgt  cccatgctgt  GGGGGctgtA
      351  CGATGTGCAT  gctcatcgcc  gACACGGCA  CGGAATGGCT  CGACAGCGCG
      401  CTGATGCGCG  CGATGAACGT  CTCATCGCG  CGCCGATCG  CCATTGCGCG
      451  CGCCAAACTG  CTGCGCTGTA  AATCCAACT  GATGTGGCGT  TTCATGCTTG
      501  CGCACAACT  GCCGACTGCG  AGCAAAATGA  TTGCCGAAAT  CAGCAACGCG
      551  AGGCGTATGA  CGCGCGAAG  TTTGGAGCAG  AATATGGTCA  AAATGCGCCA
      601  AATCAACGCA  CGCATGGTCA  AAAGCCGCG  CACCTCGGCC  GCCACATCGG
      651  GCGAAGCCG  CATCAGCCCC  TCCATGATGG  AAGCATGCA  GCACGCCAC
      701  CGCAAAATCG  TCAACACCAC  CGAGCTGCTC  CTGACCACCG  CGGCCAAGCT
      751  GCAATCTCCC  AAACCTCAAG  GCAGCGAAAT  CCGGCTGCTC  GACCGCACT
      801  TCACACTGCT  CCAACCGCAC  CTGCACACCA  CCGCGCCCT  CATCAGCG
      851  AGACACGCCC  GCCGCTACCG  CATGCACACC  GCGACTCAGC  CGCACTGGA
      901  AGCCCTCCG  GACACCTCC  ACTACCATCT  CGAGGCTTC  CTCTGGCTCA
      951  GCACCAATAT  GCGTCAGGAA  ATTTCGCGAT  TCCTATCCT  GCTGCAACGC
      1001  ACCCGCGCA  AATGGCTGGA  TGCCACGAA  GSCCACACCC  TGCGCCAAAG
      1051  CCTGCTTGAA  ACACGGGAAC  ACGGCTGA
      1101

```

This corresponds to the amino acid sequence <SEQ ID 638; ORF146ng-1>:

```

45      1  MNSGQRKRLS  GRWLNSYERY  RHRLIHAVR  LGGTVLFA  LARLLHLQH
      51  EWIGMTVFV  LGMLQFQGA  YSNAVERMLG  TVIGLGAGLG  VLWLNQHYFH
      101  GNLLFYLTIG  TASALAGWAA  VGKNYVPM  AGLTMCMLIG  DNGSEWLDG
      151  LMAMNVLIG  AATAIAAKL  LPLKSTLMWR  FMLADNLAD  SKMIAEISNG
      201  RMRTRERLEQ  NMVKMRQINA  RMVKRSHLA  ATSGESRISP  SMMEAMQHHA
      251  RKIVNTTELL  LTAAKLQSP  KLNGSEIRLL  DRHFTLLQTD  LQQTAAALING
      301  RHARRIRIDT  AINPELEALA  EHLHYQWQGF  LWLSTNMRQE  ISALVILLQR
      351  TRRKWLDAHE  RQHRLRQSLLE  TREHG*

```

ORF146ng-1 and ORF146-1 show 96.5% identity in 375 aa overlap

```

55      orf146-1.pep  MNTSQRNRLVSRWLNSYERYRRLIHAVRLGGAVLFA  TASARLLHLQHGEWIGMTVFV
      orf146ng-1    |||
      MNSGQRKRLSGRWLNSYERYRHRLIHAVRLGTVLFA  FATALARLLHLQHGEWIGMTVFV
      orf146-1.pep  LGMLQFQGAISKAVERMLGTVIGLGAGLGVLWLNQHYFHNLLFYLTIGTASALAGWAA
      |||
      orf146ng-1    LGMLQFQGAISNAVERMLGTVIGLGAGLGVLWLNQHYFHNLLFYLTIGTASALAGWAA

```

60

```

orf146-1.pep  VGKNGYVFMLAGLTMCLIGDNGSEWLDGSLMRAMNVLIGAAITAIAAKLLPLKSTIMWR
orf146ng-1     VGKNGYVFMLAGLTMCLIGDNGSEWLDGSLMRAMNVLIGAAITAIAAKLLPLKSTIMWR

5 orf146-1.pep  FMLADNLADCSKMIAEISNGRMRTRERLEENMAKMRQINARMVKSRSHLAATSGESRISP
orf146ng-1     FMLADNLADCSKMIAEISNGRMRTRERLEENMAKMRQINARMVKSRSHLAATSGESRISP

10 orf146-1.pep  AMMEAMQHAHRKIVNTTELLLTAAKLQSPKINGSEIRLLDRHFTLLQTDLQQTVALING
orf146ng-1     SMMEAMQHAHRKIVNTTELLLTAAKLQSPKINGSEIRLLDRHFTLLQTDLQQTAAALING

15 orf146-1.pep  RHARRIRIDTAINPELEALASHLHYQWQGFVLWSTNMQRQEISALVILLQRTTRKWLDAHE
orf146ng-1     RHARRIRIDTAINPELEALASHLHYQWQGFVLWSTNMQRQEISALVILLQRTTRKWLDAHE

orf146-1.pep  RQHLRQSLLETREHGX
orf146ng-1     RQHLRQSLLETREHGX

```

20 Furthermore, ORF146ng-I shows homology with a hypothetical *E.coli* protein:

```

sp|P33011|YEEA_ECOLI_HYPOTHETICAL_40.0_KD_PROTEIN_IN_COBU-SMC_INTERGENIC_REGION
>gi|1736674|gn|l|PID|d|016553 (D90838) ORF ID: o348#20; similar to [SwissProt
Accession Number P33011] [Escherichia coli] >gi|1736682|gn|l|PID|d|016560 (D90839)
ORF ID: o348#20; similar to [SwissProt Accession Number P33011] [Escherichia coli]
25 >gi|1788318 (AE000292) f352; 100% identical to fragment YEEA_ECOLI SW: P33011 but
has 203 additional C-terminal residues [Escherichia coli] Length = 352
Score = 109 bits (271), Expect = 2e-23
Identities = 89/347 (25%), Positives = 150/347 (42%), Gaps = 21/347 (6%)

30 Query: 20 YRHRRLIHAVRLGTVLFATALARLLHLQHGEWIGMTVFVVLGMLQFOGAIYSNAVERML 79
YRH R++H R+ L + RL + W +T+ V++G + F+G + A ER+
Sbjct: 15 YRHYRIVHGRVALAFLLTFLIIRLFTIFESTWPLVMTVMVIMGISFVGNVVPRAFRER 74

35 Query: 80 GTVIGLGAGLGVLWLNQHYFHGNLLFYLFIGTASALAGWAAVKNGYVFMLAGLTMCLMI 139
GTV+G GL L L L + A L GW A+GK Y +L G+T+ +++
Sbjct: 75 GTVLGSLGLIALQLE---LISLPLIMLVCAAMPFCGWLALGKKPYQGLLIGVTLAIV 131

40 Query: 140 DNGSEWLDGSLMRAMNVLIGXXXXXXXKLLPLKSTIMWRFMLADNLADCSKMIAEISN 199
G E +D+ L R+ +V++G + P ++ WR LA +L + +++ +
Sbjct: 132 GSPTGE-IDTALWRSQDVLGSLAMLFTGWPQRAFIHWRQLAKSLTEYNVVSQAFS 190

45 Query: 200 GRMTRERLEQNVMKMRQINARMVKSRSHLAATSGESRISFMSMEAMQHAHRKIVNXXXX 259
+ R RLE ++ K+ VK R +A S E+RI S+E +Q +R +V+
Sbjct: 191 FNLLEERLESHLQKLL---TDAVKMRGLIAPASKETRIPKSIYEGITINRNLVCMLEL 247

50 Query: 260 XXXXXXXXQSPK---INGSEIRLLDRHFTXXXXXXXAALINGRHARRIRIDTAINPEL 316
+ LN +R D + AL G +N +
Sbjct: 248 QINAYWATPESHEVLLNAQKLR--DTQHMQQILLSLVHALYEGNPPQPVFANTEKLNDAV 305

Query: 317 EALAEHL--HYOWQ-----GFLWSTNMQRQEISALVILLQRTTRK 354
E L + L H+ + G++WL+ ++ L L+ R RK
Sbjct: 306 EELRQLLNHHDLKVVETPIYGVLNMTETAHQLELLSNLICRALRK 352

```

On the basis of this analysis, including the identification of several transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 76

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 639>

```

1 ..GCCGAAGACA CGCGCGTTAC CGCACAGCTT TTGAGCGCGT ACGGCATTCA
51 GGGCAAACTC GTACATGTGC GCGAACACAA CGAAGCGCAG ATGGCGGACA

```

101 AGATTGTCGG CTATCTTTCA GACGGCATGG TTGTGGCACA GGTTCOGAT
 151 GCGGGTACGC CGGCCGTGTG CGACCCGGGC SCGAAACTCG CCGCCGCGGT
 201 GCGTGAGGCC GGGTTTAAAG TCGTCCCGT CGTGGCGCA AC, GGGTGA
 251 TGCGGCTTT GACGCTGGC GCGTGGGAA AGTCGCGATT TTATTCTAAC
 301 GGTTTTCTTAC CGCCGAAATC GCGAGAACGC AGSAACTGT TTGCCAAATG
 351 GGTGCGGGCG CGCTTTCCTA TCGTCATGTT TGAACGCCCG CACCGCATCG
 401 GTGCAGCGCT TCCGATATG GCGGAACGTG TCCCGAACG CCGATTATG
 451 CTGGCGCGCG AAATTACGAA AACGTTTGAA ACCTTCTTAA CGGCGACGGT
 501 TGGGGAAATT CAGACGGCAT TGTCTGCCGA GCGGACCAA TCGCGCGCG
 551 AGATGGTGTT GGTGCTTTAT CCGCGCAGG ATGAAAAACA CGAAGCCGTTG
 601 TCCGAGTCCG CGCAAAACAT CATGAAATC CTCACGCCG AGCTGCCGAC
 651 CAACAGCGCG GGGAGCTTG CTGCCAAAT CACGGCGGAG GGAAGAAAG
 701 CTTTGTACGA T. .

This corresponds to the amino acid sequence <SEQ ID 640; ORF147>:

15 1 ..AEDTRVTAQL LSAYGIQGLK VSVREHNERQ MADKIVGYLS DGMVVAQVSD
 51 AGTPAVCDPG AKLARRVREA GFKVVPVGA XAVMAALSA VGEVSDFYFN
 101 GFVPPKSGER RKLFAKWVA AFPIVMFETP HRIGAALADM AELFPERRIM
 151 LAREITKTFE TFLSGTVEI QTLASADGQ SRGEMVLVLY PAQDEKHEGL
 201 SESAQINIKI LTAELPTQQA AELAAKITGE GKALYD. .

20 Further work revealed the complete nucleotide sequence <SEQ ID 641>:

1 ATGTTTCAGA AACATTGCA GAAAGCTCC GACAGCGTC TCGAGGGAC
 51 ATTATACGTG GTTGCCACGC CCATCGGCAA TTGCGCGCAT ATTACCTGTC
 101 GCGCTTTGGC GGTATTGCAA AAGCGGACA TCATCTGTGC CGAAGACACG
 151 CGCGTTACGC CACAGCTTTT GAGCGGTAC GCAACTCAGT GCNAACTCGT
 201 CAGTGTGCGC GAACACACGC AACGCGCAGT GCGGACAGG ATTGTCCGCT
 251 ATCTTTTACA GCGCATGGT TTGGCACAGG TTCCGATGC GGGTACGCGC
 301 GCCGTGTGCG ACCCGGGCGC GAAACTCGCC CGCCGCGTGC GTGAGGCGCG
 351 GTTTAAAGTC GTTCCGTCG TGGCGCAAG CGCGGTGATG CGCGCTTTGA
 401 GCGTGGCCGG TGTGGAAGGA TCCGATTTT ATTTCAAAGG TTTTGTACCG
 451 CGAANAATCG GAGAACGCGA GAACTGTTT GCCAATGGG TCGCGCGCGC
 501 GTTTCCTATC GTCATGTTTG AAACGCGCGA CCGCATCGGT GCGACGCTTG
 551 CGGATATGCG GCAACTGTTC CCGACACGCC GATTAATCGT GCGCGCGGAA
 601 ATTACGAAAG CTTTGAACG GTTCTTAAGC GGCACGCTTG GGAATATCA
 651 GACGGCATTC TCTGCCGACG GCAACCAATC AAGGCTTGTG CAGCTCCGCG
 701 TGCTTTATCC GCGCGCAGAT GAAAACACG AAGGCTTGTG CAGCTCCGCG
 751 CAAACATCA TGAATACTCT CACAGCCGAG CTGCCGACCA AACAGCGCGC
 801 GGAGCTTGCT GCCAAATCA CCGGCGAGGG AAGAAAGCT TTGTACGATC
 851 TGGCTCTGTC TTGAAAAAC AAATAG

This corresponds to the amino acid sequence <SEQ ID 642; ORF147-1>:

40 1 MFQKHLQKAS DSVVGSTLYV VATPIGNLAD ITRLALAVLQ KADIICAEDT
 51 RVTAQLLSAY GIQGLVSVR EHNERQMAK IVGYLSDGMV VAQVSDAGTP
 101 AVCDFGAKLA RRVREAGFKV VPVVGASVM AALSVAAGVE SDFYFNGFVP
 151 PKSGERRKLF AKWVRAAFPI VMFETPHRIG ATLADMAELF PERRIMLARE
 201 ITKTFTFLS GTVGEIQTAL SADGNQSRGE MVLVLYPAQD KEHGLSESA
 45 251 QNIMKILTA LPTKQAELA AKITGEGKKA LYDLALSKM K*

Computer analysis of this amino acid sequence gave the following results:

Homology with hypothetical protein ORF286 of *E. coli* (accession number U18997)

ORF147 and *E. coli* ORF286 protein show 36% aa identity in 237aa overlap:

50 Orf147: 1 AEDTRVTAQLLSAYGIQGLVSVREHNERQMAKIVGYLSDGMVVAQVSDAGTPAVCDPG 60
 AEDTR T LL +GI +L ++ +HNE+Q A+ ++ L +G +A VSDAGTP + DPG
 Orf286: 43 AEDTRHTGLLLQHFGINARLFAHDHNEQQKAETLLAKLQEQGNLALVSDAGTPLINDPG 102
 Orf147: 61 AKLARRVREKXXXXXXXXXXXXXXXXXXXXXGSDFYFNGFVPKSGERRKLFAKWVA 120
 L R RE F + GF+P KS RR
 55 Orf286: 103 YHLVRTCEAGIRVVPLPGPCAITALSAAGLPSDRFCYEGFLPAKSKGRDRDALKAEAE 162
 Orf147: 121 AFPIVMFETPHRIGAALADMAELFPERR-LMLAREITKTFELSGTVEIQTALASADGD 179
 ++ +E+ HR+ +L D+ + ER ++LARE+TKT+ET VGE+ + D+
 Orf286: 163 PRTLIFYESTHRLDLSLEIDIVVLGESRYVVLARELTKTWTETIHGAPVGELLAWVKEDEN 222

Orf147: 180 QSRGEMVLVLPAQDEKHEGLSESAQNIMKILTAELPTKQAAELAAKITGEGKKALY 236
+ +GEMVL++ + E L A + +L AELP K+AA LAA+I G K ALY
Orf286: 223 RRGKEMVLIV-EGHKAQEEDLPADALRTLALLQAEPLKKAALAAEIHGVKKNALY 278

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF147 shows 96.6% identity over a 237aa overlap with ORF75a from strain A of *N. meningitidis*:

```

                                     10          20          30
orf147.pep      AEDTRVTAQLLSAYGIQCKLIVSVREHNERQ
                                     |||
orf75a      TLYVYVATPIGNLADITLRALAVLKADIICAEDTRVTAQLLSAYGIQCKLIVSVREHNERQ
               20          30          40          50          60          70          80          90
orf147.pep      MADKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLARVRREAGFKVVPVVGAXVMAALSVA
               |||
orf75a      MADKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLARVRREVGFKVVPVVGASAVMAALSVA
               80          90          100         110         120         130
orf147.pep      GVGESDFYFNGFVPPKSGERRLFAKWVRAAFPVFMFETPHRIGAALADMAELFFERRIM
               |||
orf75a      GVGESDFYFNGFVPPKSGERRLFAKWVVRVAFPVVMFETPHRIGATLADMAELFFERRIM
               140         150         160         170         180         190
orf147.pep      LAREITKTFTFSLSGTVGIEQTALSADGQDSRGEMVLVLYPAQDEKHEGSESQAQNMIMKI
               |||
orf75a      LAREITKTFTFSLSGTVGIEQTALAADGNQDSRGEMVLVLYPAQDEKHEGSESQAQNMIMKI
               200         210         220         230         240         250
orf147.pep      LTAEPLTKQAELAARKITCGEKKALYD
               |||
orf75a      LTAEPLTKQAELAARKITCGEKKALYDLALSWNKX
               260         270         280         290

```

ORF147a is identical to ORF75a, which includes aa 56-292 of ORF75.

Homology with a predicted ORF from *N.gonorrhoeae*

ORF147 shows 94.1% identity over a 237aa overlap with a predicted ORF (ORF147ng) from *N.*

gonorrhoeae:

orf147.pep		AEDTRVTAQLLSAYGQGGKLVSVSRHNEHQ	30
orf147.ng	TLVVVATPIGNLADITRLALVLQKADI	CAEDTRVTAQLLSAYGQGGKLVSVSRHNEHQ	85
orf147.pep	MADKIVGGLSDGMVVAQVSDGATPAVCDPGAKLARVR	REAGFKVVPVVGAXAVMAALSVA	90
orf147.ng	MADKIVGLSDGLVVAQVSDGATPAVCDPGAKLARVR	REAGFKVVPVVGASAVMAALSVA	145
orf147.pep	GVGSDSYFYNGFVPPKSGERRKLFAKWVRAAFP	IVMFETPHRIGAAALDMAELFPERRLM	150
orf147.ng	GVASDSFYNGFVPPKSGERRKLFAKWVRAAFP	VMVFETPHRIGATLADMAELFPERRLM	205
orf147.pep	LAREITKTFETFLSGTVEIGIQTALSDAGDSQRGEM	VLVLPQADEKHGELSSEAGNIHKI	210
orf147.ng	LAREITKTFETFLSGTVEIGIQTALADGDSQRGEM	VLVLPQADEKHGELSSESAGNMKI	265
orf147.pep	LTAEPLTKQAELAELAAKITGEGKKALYD	237	
orf147.ng	LAELPLTKQAELAELAAKITGEGKKALYDLALSWKNK	300	

An ORF147ng nucleotide sequence <SEQ ID 643> was predicted to encode a protein having amino acid sequence <SEQ ID 644>:

```

1 MSVFQTAFFM FQKHLQKASD SVVGGTLYVYV ATPIGNLADI TLRALAVLQK
51 ADIIICAEDTR VTAQLLSAYG IQGRLVSVRE HNERQMADKV IGFLSDGLVYV
101 AQVSDAGTPA VCDPGAKLAR RVREAGFKVY EVVGASAVMA ALSVAGVAES
151 DFYFNGFVFP KSGERRKLFA KWVRAAFVYV MFETPHRIGA TLADMAELFP
201 ERLRLMLAREI TKTFTFLSG TVGIEQTALA ADGNQSRGEM VLVLYPAQDE
251 KHEGLSESAQ NAMKILAAEL PTKQAELAA KITGEGKKAL YDLALSWKNK
301 *

```

10 Further work revealed the following gonococcal DNA sequence <SEQ ID 645>:

```

1 ATGTTTCAGA AACACTTGCA GAAAGCCTCC GACAGCGTCG TCGSAGGAC
51 ATTATACGTG GTTGCCACGC CCATCGGCAG TTTGCCAGAC ATTACCGCTG
101 GCGCTTTGCG GGTATTGCGA AAGCGGACAC TCATTGTGCG CGAAGACACG
151 CGCGTTACTG CGCAGCTTTT GAGCGCGTAT GGCATTTCAG GCAGGTTGGT
201 CAGTGTGCGG GACACACACG AGCGGCAGAT GCGCGACRAG GTAATCGGTT
251 TCCTTTTCAG CGGCTTGGTT GTGGCGCAGG TTTCGAGTGC GGGTAGCCGC
301 GCGGTGTGCG ACCCGGGCGC GAAACTCGCC CGCCCGGTGC GCGAAGCAGG
351 GTTCAAAGTC GTTCCGCTCG TGGCGCAGAG CGCGTAAATG GCGGCGTTGA
401 GTGTGGCCGG TGTGGCGGAA TCCGATTTTT ATTTCAACGG TTTTGATCCG
20 451 CGGAATTCGG GCGAAGCTAG GAAATTTGTT GCCAATGGG TCGGGGCGGC
501 ATTTCTCTGT GTCATGTTTG AAACCGCGCA CGGAATCGGG GCAACGCTTG
551 CGGATATGCG GGAATTTGTC CCCGAACGCC GTCTGATGCT GCGCGCGGAA
601 ATCACGAAAA CGTTTGAACG GTTCTTAAGC GCGACGTTTG GGGAAATTC
25 651 GACGCGATTG GCGGCGGACG GCAACCAATC GCGCGCGGAG ATGGTGTGTTG
701 TGCTTTATCT GCGCGAGGAT GAAAAACACG AAGGCTTGTC CGAGCTTCGG
751 CAAATGCGA TGAATATCCT TGCGGCGGAG CTGCGGACCA AGCAGCGCGC
801 GGAGCTTGCC GCAAGATTA CAGGTGAGGG CAAAAGGCT TTGTACGATT
851 TGGCACTGTC GTGAAAAAAC AATATGA

```

This corresponds to the amino acid sequence <SEQ ID 646; ORF147ng-1>:

```

30 1 MFQKHLQKAS DSVVGGTLYV VATPIGNLAD ITLRALAVLQ KADIIICAEDT
51 RVTAQLLSAY GIQGRLVSVR EHNERQMADK VIGFLSDGLV VAQVSDAGTP
101 AVCDPGAKLA RRVREAGFKV EVVGASAVM AALSVAGVAE SDFYFNGFVP
151 PKSGERRKLF AKWVRAAFVY VMFETPHRIG ATLADMAELF PERRRLMLARE
201 ITKTFTFLSG GTVGEIQTAL AADGNQSRGE MVLVLYPAQD EKHEGLSESA
35 251 QNAMKILAAE LPTKQAELAA AKITGEGKKA LYDLALSWKN K*

```

ORF147ng shows homology to a hypothetical *E. coli* protein:

```

sp|P45528|YRAL_ECOLI HYPOTHETICAL 31.3 K D PROTEIN IN AGAI-MTR INTERGENIC REGION
(F286)
>gi|606086 (U18997) ORF_f286 [Escherichia coli]
40 >gi|1789535 (AE000395) hypothetical 31.3 kD protein in agai-mtr intergenic region
[Escherichia coli] Length = 286
Score = 218 bits (550), Expect = 3e-56
Identities = 128/284 (45%), Positives = 171/284 (60%), Gaps = 4/284 (1%)

45 Query: 4 KHLQKASDSVVGTTLYVATPIGNLADITLRALAVLQKADIIICAEDTRVTAQLLSAYGIQ 63
K Q A +S G LY+V TPIGNLADIT RAL VLQ D+I AEDTR T LL +GI
Sbjct: 2 KQHQSADNSQ--GQLYIVTPTIGNLADITQRALEVLQAVDLIAEDTRHTGLLLQHFGIN 59

50 Query: 64 KRLVSVREHNERQMADKVIIGFLSDGLVVAQVSDAGTPAVCDPGAKLARVREAGFKVFPV 123
RL ++ +HNE+Q A ++ L +G +A VSDAGTP + DFG L R REAG +VVP+
Sbjct: 60 ARLFALDHNEQQKAEITLLAKIQEGONIALVSDAGTPLINDPGYHLVTRCAGIRVVF 119

55 Query: 124 VGASAVMAALSVAGVAESDFYFNGFVPPKSGERRKLFKAWVRAAFVVMFETPHRIGATL 183
G A + ALS AG+ F + GF+P KS RR ++ +E+ HR+ +L
Sbjct: 120 PGPCAITALSAAGLPSDRFCYEGFLPAKSGRGEIDALKATEAPETLIFYESTHRLDLS 179

60 Query: 184 ADMAELFPERR-LMLAREITKTFTFLSGTVGIEQTALAADGNQSRGEMVLVLYPAQDEK 242
D+ + E R++LARE+TKT+ET VGE+ + D N+ +GEMVL++ +
Sbjct: 180 EDIVAVLGESRYVVLARELTKTWTETIHGAPVGEILLAWKEDENRRKGMVLV-EGHKAQ 238

60 Query: 243 HEGLSESAQNAMKILAAELPTKQAELAAKITGEGKKALYDLAL 286

```

E L A + +L AELP K+AA LAA+I G K ALY AL
 Sbjct: 239 EEDLPADALRTLALLQAEFLPKKAAALAAEIHGVKKNALYKYAL 282

Based on the computer analysis and the presence of a putative transmembrane domain in the gonococcal protein, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 77

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 647>

```

10 1 ATGAAACAA CCGACAAAG GACAACCGAA ACACACCGCA AAGCCCCGAA
51 AACCAGTTCG ATCCGCTTCT C.GCTGCTTA CTTAGCCATA TGCCTTCGCT
101 TCGGCATTCT TCCCCAGGCG TGGGCGGGAC ACACCTATTAT CGGCATCAAC
151 TACCAATACT ATCGCGACTT TGCCGAAAT ATAGCGAAGT TTGCAGTCGG
201 GCGCAAGAGT ATTGAGGTTT ACAACAAAAA AGGGGAGTTG GTCGGCAAT
251 CAATGACAAA AGCCCCGATG ATTGATTTTT CTGTGGTGTG GCGTAACGGC
301 GTGGCGGcAT TGGTGGGCGT ATCAATATAT TGTGAGCGTG GCACATAACG
15 351 GCGGCTATAA CAACGTTGAT TTTGTGCGGG AAGGAAR. AA EATCCC. SAT
401 CAACAwCGwW TTACTTTATAA AATTGTGAAA CGGAATTAAT ATAAAGCAG
451 GACTAAAGCG CATCCTTATG CGCGCGAITA TATATTCGCG COTTTCGATA
501 AATWGTGCAC AGATGCGAAA CCTGTGAAA TGACACAGTTA TATGGATGGG
551 CGGAATATA TCGATCAAAA TAATTACCGT GACCGTGTTC GTATTGGGGC
20 601 ASGCAGGCAA TATTGCCGAT CTGATGAAGA TGAGCCCAAT AACCGCAAAA
651 GTTCATATCA TATTGCAAGT .....
701 .....GCCTC ACCAATGTTT ATCTATGATG CCGAAAAAGCA
751 AAATGGTTTA ATTAATGGGG TATTGCAACG GGGCAACCCC TATATAGGAA
801 AAAGCAATGG CTTCAGCTG GTTCGTAAAG ATTGTTCTA TGAATGAATC
25 851 TTTGCTGGAG ATACCCATTC AGTATTCTAC GAACACAGTC AAAATGGGAA
901 ATACTCTTTT AACGACGATA ATAATTGGCA AGGA AAAATC AATGCCAAC
951 ATGAACACAA TTCTCTGCCT AATAGATTAA AACACGAAC CGTTCAAATT
1001 TTTAATGTTT CTTTATCCGA GACAGCAAGA GAACCTGTTT ATCATGCTGC
30 1051 AGGTGGTGTG AACAGTTATC GACCCAGACT GAATATATGA GAAATATTT
1101 CTTTATTTGA CGAAGGAAAA GCGGAATTGA TACTTACCAG CAACATCAAT
1151 CAAGGTGCTG GAGGATTATA TTTCCAAGGA GATTTTACGG TCTCGCTGTA
1201 AATAAACGAA ACTTGGCAAG GCGCGGGCGT CTATATCAGT GAAGACATA
1251 CCGTTACTTG GAAAGTAAAC GCGGTGGCAA ACACACCGCT GTCCAAATC
35 1301 GGCAAGGCAA CGCTG. ....
//
2101 ..... ..GATAAAG
2151 TGACTGCTTC ATTGACTAAG ACCGACATCA GCGGCAATGT CGATCTTGGC
2201 GATCAGCGTC ATTTAAATCT CACAGGGCTT GCCACATCA ACGGCAATCT
2251 TAGTGCAAAAT GCGGATACAC GTTATACAGT CAGCCACAAC GCCACCAAAA
40 2301 ACGGCAACCK TagCctCgtG G. sAATGcCC AAGCAACATT TAATCAAGCC
2351 ACATTAAACG GCAACACATC GCCTTCgGGC AATGCTTCAT TTAATCTAAG
2401 CGACCACGCC GTACAAAACG GCAGTCTGAC GCTTTCGGCG AACGCTAAGG
2451 CAAACGTAAG CATTTCGCGA CTCACCGGTA ATGTCTCCCT AGCGCATTAAG
45 2501 CGAGTATTCC ATTTTGAAAG CAGCGGCTTT ACCGGACAAA TCGAGCGGGCG
2551 CAagGATACG GCATTACACT TAAAGACAGC CGAATGGACG CTGCGCTCag
2601 GarCGGAATT AGGCAATTTA AACCTTGACA ACGCCACCAT TACACTCAAT
2651 TCGCCTATC GGCACGATTG GCGAGGGCGC CAAACCGGCA GTGGGACGA
2701 TGGCGCGCGC CGCGGTTCG CACCGTTCCTA TTAATGCTGA
50 2751 CACCGCACC TTGCGTGAAT TCCTGTTTCA ACACGCTGAC GGTAAACGGC
2801 AATTCAACG GTCAGCGAAC ATTCGCGTTT ATGTGCGAAG TCTTCGCTGA
2851 CCGCAGCGAC AATTGGAAC TGGCGGAAG TTCGGAAGC ACTTACACCT
2901 TCGCGGCTCA CAATACCGCG AACGAACCTG CAAGCTCGA ACAATTGAGC
2951 GTAGTGAAG GAAAGACAA CAACCGCTG TCCGAAACCC TTAATTTCAC
3001 CTTGCAAAAC GAACACGTCG ATGACGGCGC GTGG. ....
//
3551 ..... ..TTAGAC GCGGTATTGT CCGAAGACCG
3601 CGCAGACGCC GTTTGGACAA GCGGCATCCG GGACACCAAA CACTACCGTT
3651 CGCAAGATTT CGCGCCTTAC GCACAACAAA CCGACCTGGC CCAATCTGGT
70 3701 ATGACAGAAA ACCTCGGCAG CGGGCGGCTG GGCATCTGT TTTGACAAA
3751 CCGGACCGAA AACACCTTCG ACAGCGGATC CGGCAACTCG GCACGGCTTG
3801 CCCACGGCGC GGTTTTCGGG CAATACGGCA TCGACAGGTT CTACATCGGC

```

	3851	ATCAG ₁₁ CGCG	GCGCGGGGTT	TTAGCAGCGG	CAGCGCTTTcA	GACGGCATCG
	3901	GAGsmAAAwT	CGCGCGCGCG	GTGCTGCATT	ACGGCATTCA	GGCACGATAC
	3951	CGCGCGGGtt	tCgGCGgAtt	CGCGATCGAA	CGGCACATCG	GGCCAACGcg
5	4001	cTATTTCGTC	CAAAAACGCG	ATTACCGCTA	CGAAAACGTC	AATATCGCA
	4051	CCCGCGGCT	TCGATTCAC	CGcTACCGCG	CGCGCATTA	GCGAGATTAT
	4101	TCATTCGAAC	CGCGCGAACA	CATTTCcCAT	ACGGCTTATT	TGAGCGCTGC
	4151	CTATACCGAT	CGCGCTTCGG	GCAAGTCCG	AACACGGCTC	AATACCGCGC
	4201	TATTGGCTCA	GGATTTCGGC	AAAACCGCA	GTGCGGAATG	GgGCGTAAC
10	4251	GCGGAAATCA	AAGGTTTCAC	GCTGTCCCT	CACGCTGCCG	CGCGCAAAGG
	4301	CGCGCAACTG	GAAAGCGCAAC	ACAGCGCGGG	CATCAAAATA	GGCTACCCGCT
	4351	GGTAA...				

This corresponds to the amino acid sequence <SEQ ID 648; ORF1>:

	1	MKTTDKRTTE	THRKAPKTGR	IRFXAAYLAI	CLSFGLPQA	WAGHTYFGIN
	51	YQYRDFAEEN	KGKFAVGAKD	IEVYNKKGEL	VGKSMTKAPM	IDFSVVSRRK
15	101	VAALVGVQYI	VSVAHNGGYN	NVDFAEGXND	IXDQXRXITYK	IVKRNRYKAG
	151	TKGHFVGGDY	HMPRLHKVYT	DAEPVEMTSY	MDGRKYIDQN	NYPDPRVRGA
	201	GRQYWRSDDE	EPNNRESSYH	IAS.....GS	PMFIYDAQKQ
	251	KWLINGVLQT	GNPYIGKSNQ	QFLVRKDWFY	DEIFAGDTHS	VFEYEPQNGK
	301	YSFNDDNNGT	GRINAKHEHN	SLPNRLKTRT	VLQNVLSLE	TAREFVYHAA
20	351	GGVNSYFRRL	NGENISFID	EGKGELITLS	NINGGAGGLY	FQGDFTVSPE
	401	NNETWQAGVG	HISEDSTVTW	KVNGVANRDL	SKIGKGLT...
				//		
	701DKVTAS	LTKTIDISGNV	DLADHAHLNL	TGLATLNGNL
	751	SANGDTRYTV	SHNATONGNX	SLVXNAQATF	NQATLNGNTS	ASGNASFNLS
25	801	DHAVONGSLT	LSGNANANVS	HSALNGNVSL	ADKAVFHEFS	SRFTQGISGG
	851	KDTALHLKDS	EWTLPFGKEL	GNLNLDNATI	TLNSAYRHDA	AGAQTGSATD
	901	APRRRSRRSR	RSLLXVTPEP	SVESRFTNLT	VNGKLNGQGT	FRFMSELFYG
	951	RSDKLKLAEIS	SEGTYTILAVN	MTGNEFASLE	QLTVVEGKDN	KPLSENLMFT
30	1001	LQNEHVDAQA	W.....
				//		
	1151LDRVFAEDR
	1201	RNAVWTSQIR	DTKHYSRQDF	RAYRQQTDLR	QIGMQKNLGS	GRVGLIFSHN
	1251	RTENTPDGDI	ONSARILANG	VFGQYGDIFR	YIGISAGAGF	SSGSLSDGIG
35	1301	KXRRVVLHY	CTAARVAFAT	CGFGIBPHIG	ATRYFVQKAD	YRYENVNIAI
	1351	PGLAFNRYRA	GKADYSEKQ	GLSISITPYIL	SLSYTDARAS	KVRIRNVAIV
	1401	LADQFCKTRS	AEWGVNAEIK	GFTLSLHAAA	AKGFPQLAEQH	SAGIKLQFWR
	1451	*				

Further sequencing analysis revealed the complete nucleotide sequence <SEQ ID 649>:

	1	ATGAAACCAA	CGACAAACG	GACAACCGAA	ACACACCGCA	AAGCCCGGAA
40	51	AACCGCGCGC	ATCCGCTTCT	CGCCTGCTTA	CTTAGCCATA	TGCCTGTGGT
	101	TCGGCATCTC	TCCCAAGGCC	TGGCGGGAGC	ACACTTATT	GGCGATCAAC
	151	TACCAATACT	ATCGCGACTT	TGCCGAAAAA	AAAGGCAAGT	TTCAGTCCGG
	201	GGCGAAAGAT	ATTGAGGTTT	ACACACAAAA	AGGGGAGTTG	GTCCGCAATC
45	251	CAATGACAAA	AGCCCCGATG	ATTGATTTTT	CTGTGGTCTG	GGGTAAACCG
	301	GTGCGCGCAT	TGGTGGGCGA	TCAATATATT	AGGAGCGTGG	CACATAACCG
	351	CGGCTATAC	AACTGTGATT	TGGTGGCGGA	AGGAGAGAAAT	CCCGATCAAC
	401	ATCGTTTATC	TTTATAAATC	GTGAACACGA	ATAATTATAT	AGCAGGGAAT
	451	AAAGGCCATC	CTTATGCGCG	CGATATATAT	ATGCGCGCTT	TGCATTAART
50	501	TGTACACAGT	CGACAAACCT	TTGAATATAC	CAGTTATATG	GATGGCGGCA
	551	AATATATCGA	TCAAAATAAT	TACCCGTACC	GTGTTTCGAT	TGGGCGAGCG
	601	AGGCAATATT	GGCGATCTGA	TGAAGATGAG	CCCAATAACC	CGGAAAGTTC
	651	ATATCATATT	CGAAGTTCGT	ATTCTTGGCT	CGTTGGTGGC	AATACCTTTG
	701	CACAAATATG	ATCAGTGGGT	GGCAGATGCA	ACTTAGGTAG	TGAAAAAATT
55	751	AAACATAGCC	CATATGGTTT	TTTACCAACA	GGAGGCTCAT	TTGCGCAGAG
	801	TGGCTCACCA	ATGTTTATCT	ATGATGCCCA	AAAGCAAAGG	TGGTTAATTA
	851	ATGGGGTATT	GAAAACGGCG	AAOCCCTATA	TAGGAAATAAC	CAATGGCTTC
	901	CAGCTGGTTC	GTAAGAGATT	GTTCATGAT	GAAATCTTTG	CTGGAGATAC
	951	CGATTTCAGTA	TTCACGAAC	CAGCTCAAAA	TGGGAAATAC	TCCTTTAAGG
60	1001	ACGATATATTA	TGCACAGAGA	AAATCAATG	CCAAACATGA	ACCAATTTCT
	1051	CTGCCTAATA	GATTAAARAC	AGCATTCAGT	CARTTCTTTA	ATGTTTCTTT
	1101	ATCCGAGACA	CGAAGAGAAC	CTGTTTATCA	TGCTGCAGAT	CGTCTACACA
	1151	GTTATCGACC	CAGACTGAAT	AATGAGAGAA	ATATTTCCCT	TATTGACAAA
	1201	GAAAAGGCGC	AATTGATACT	TACCGACAAC	ATCAATCAAG	GTGCTGAGG
65	1251	ATTATATTTT	CAAGGAGATT	TTACGGTCTC	GCCTGAAAT	AACGAAACTT
	1301	GGCAAGGCGC	GGGCGTTCAT	ATCAGTGAAG	ACAGTACCGT	TACTTGAGAA
	1351	GTAACCGCGC	TGGCAACAGA	CCGCTCTGCC	AAATCGGCA	AAGGACGCTT

1401 GCACGTTCAA GCCAAGGGG AAAACCAAGG CTGCATCAGC GTGGGCGAGC
 1451 GTACAGTCAT TTCCGATCAG CAGGCAGACG ATTAAGGCGCAA AAAACAGGCC
 1501 TTACTGAAA TCGGCTTGGT CAGCGGCAGG GGTACGGTGC AACTGAATGC
 1551 CGATAATCAG TTCAACCCCG ACAAACTCTA TTTCCGGCTT CGCGCGCGGAC
 1601 GTTTGGATTT AAACGGGCAT TCGCTTTTCG TCCACCGTAT TCAAAATACC
 1651 GATGAAGGGG CGATGATTGT CAACACAAT CAAGACAAAG AATCCACCGT
 1701 TACCATTACA GGCATAAAG ATATTGCTAC AACCGGCAAT AACACAACGT
 1751 TGGATAGCAA AAAAGAAATT GCCTACAACG GTTGGTTTGG CGAGAAGATG
 1801 ACGACCAAAA CGAACGGGCG GCTCAACCTT GTTTACCAAG CCGCCGCGAG
 1851 AGACCGCACC CTGCTGCTTT CGGCGGGAAC AAATTATTAAC GGCAACATCA
 1901 CGCAACAAAA CGGCAAACTG TTTTTCAGCG CAGAGCAACG ACCGCAACGC
 1951 TACAATCAT TAAACGACCA TTGGTCGCAA AAGAGGGGCA TTCTCCGGG
 2001 GGAATCGTG TGGGACAACG ACTGGATCAA CCGCACATTT AAGCGGAAAA
 2051 ACTTCCAAAT TAAAGGCGGA CAGGCGGTGG TTTCGCCGAA TGTTCGCAAA
 2101 GTGAAGGGCG ATTGGCATTT GAGCAATCAC GCCCAAGCAG TTTTGGTGT
 2151 CGCACCGCAT CAAGGCCACA CAATCTGTAC ACGTTCGAGC TGGACGGCTC
 2201 TGACAAATTC TGTGAAAAA ACCATTACCG ACGATAAAGT GATCTGCTCA
 2251 TTGACTAAGA CGGACATCAG CGGCAATCTG GATCTGCGC ATCACGCTCA
 2301 TTTTAAATCT ACAGGCGCTG CCACACTCAA CGGCAATCTT AGTCCAAATG
 2351 GCGATACACG TTTATACAGT AGCCACAACG CCACCCAAAA CGGCAACCTT
 2401 AGCCTCTGTG GCAATGCCCA AGCAACATTT AATCAAGCCA CATTAAACGG
 2451 CAACACATCG GCTTCGGGCA ATGCTTCATT TAATCTAAGC GACCACGGCG
 2501 TACAAAAACG CAGTCTGACG CTTTCGGGCA ACGCTAAGGC AAACGTAAGC
 2551 CATTCCGAC TCAACGCTAA TGTCCTCCCTA GCGGATAAGG CAGTATTCCA
 2601 TTTTGAAAGC AGCCGCTTTA CCGGACAAAT CAGCGCGGCG AAGGATACGG
 2651 CATTACACTT AAAAGACAGC GAATGGACCG TGCCGTACGG CACGGAATTA
 2701 GGCATTTTAA ACCTTGACAA CGCCACCATT ACACCTCAAT CGCCTATCG
 2751 CCACGATGCG CAGGCGGCGC AAACCGGCGG TGGCAGAGT ACGCGCGGCG
 2801 GCGGTTGCGC CGGTTGCGCG CGTTCCCTAT TATCGGTTAC ACCGCGCACT
 2851 TCGGTAGAAT CCGGTTTCAA CAGCGTGAGC GTAAACGGCA AATTGAACGG
 2901 TCAGGGAACA TTCCGCTTTA TGTCCGAAGT CTTCGCGTAC CGCAGCGACA
 2951 AATTGAAGCT TCGCGAAGGT TCCGAAGACA CTTCACCTT GCGCGTCAAC
 3001 AATACCGGCA ACCGACCTGC AAGCTTCGAA CAGTTCGAGC TAATTTACCT
 3051 AAAGACACAC AACCGCTGCA CGGAACAACT TAATTTACCT CTGCAAAACG
 3101 AACAATCGC TGCAGCGGCG TGGCGTTTAC ACTCATCTCG CAAGAAGCGC
 3151 GAGTTCCGCC TGCATAATCC GGTCAAAGAA CAGAGAGCTT CCGACAACCT
 3201 CGCAGAGCGA GAAGCCAAAA AACAGCGGGA AAAGACAAAC CGCGCAAGCC
 3251 TTGACGCGCT GATTGCGGCC GGGCGCGATG CCGTCGAAAA GACAGAAAGC
 3301 GTTGCCGAAC CGGCCCGGCA GGCAGGCGGG GAAAATGTCT GCAATTTATGA
 3351 GGGCGAGGAA GAGAAAAAAC GGGTGCAGCG GGATTAAGAC ACCGCTTTGG
 3401 CGAAACAGCG CGAAGCGGAA ACCCGGCGCG GTACCAACGC CTTCGCCGCG
 3451 GCCCGCGCGC CCGCGCGGGA TTTGCGCGAA CTGCAACCCC AACCGCAGCC
 3501 CCAACCGCAG CGGACCTTGA TCAGCGGTTA TGCCAATAGC GGTTTGATG
 3551 AATTTTCCGC CAGGCTCAAC AGCGTTTTTC CGGTACAGGA CGAATTAGAC
 3601 CGGCTATTTC CGGAAGACCG CGCAACGCC TTTTGGACAA CGCGGCTCCG
 3651 GGCACCCAAA CACTACCGTT CGCAAGATTT CCGCGGCTAC CGCCACAAAA
 3701 CGGACCTCGG CCAATTCGGT ATGCAGAAAA ACCTTCGGCA CCGCGCGCTC
 3751 CGCATCTGCT TTTCGACAGA ACCGCGGAA AAGCACTTCC ACGAGGCTC
 3801 CGGCACTCG GCACGCGTTG CCCACGGCGC CGTTTTCCGG CAATACGGCA
 3851 TCGACAGGTT CTACCTCGGC ATCAGCGCGG CCGCGGGTTT TAGCGCGGCG
 3901 AGCCTTTTCA CCGGCACTCG AGGCAAAATC CGCGCGCGCG TGCTGCAATTA
 3951 CGGCATTTCG CACACGATAC CGCGCGTTTT CGCGCGGATT GGCATCGAAC
 4001 CGCACATCGG CGCAACGCGC TATTTCGTCT AAAAAGCGGA TTACCGCTAC
 4051 GAAACGTCGA ATATCGCCAC CCCCGGCTTT GCAATCAACC GCTACCGCGC
 4101 GGGCATTAAG CGAGATTATT CATTCAAAAC GGGCGCAAC ATTTCATCA
 4151 CGCCTTATTT GAGCCTGTCC TATACGATG CGCTCTCGGG CAAAGTTCGA
 4201 ACACGCGTCA ATACCGCGGT ATTGGCTCAG GATTTTCGCA AAACCCGACG
 4251 TCGGGAATGG GCGGTAACAG CGGAATCAA AGGTTTACAG CTGCTCTCC
 4301 ACGCTGCCGC CGCCAAAGGC CCGCAACTGG AAGCGCAACA CAGCGCGGGC
 4351 ATCAAAATTAG GCTACCGCTG GTAA

This corresponds to the amino acid sequence <SEQ ID 650; ORF1-I>:

1 MKTIDKRTTE THRKAPKTGR TRFSPAYLAI CLSFGILLPA WAGHTVYEGIN
 51 IQVYRDPAEN KGFPAVGAKD IEVYNKKGEL VGKSMTKAEM IDFSVVSRRG
 101 VAALVYGQDY VSVAHNGGYN NVDFEAGERN POHQRTYKI VKRNNYKAGT
 151 KGHFYGGDYH MPRLHKFVID AEPVEMTSYM DGRKYIDQNN YDPRVRIAG
 201 RQVYRSDEDE PMNRESSYHI ASAYSVLWVG NTRFQNGSGG GTVNLGSEKI
 251 KHSFYGLPLPT GSGFDGSGSP MFIYDAQKOK WLINGVLQTG NPYIKGKNGF
 301 QLVKRDWFYD EIFAGDTHSV FYEPRNGKGY SFNDNNNGTG KINAKHEHNS

351 LPNRLKTRTV QLFNVSLSET AREPVYHAAG GVNSYRPRIN NGENISFIDE
 401 KGKGLILTSN INQAGAGLYF QQDPTVSPEN NETWQAGGVH ISEDSVTWTK
 451 VNGVAMDRLS KLGKGTLLHVQ AKGENQGSIS VGDGTYYLDQ QADDKGGKQA
 551 FSEIGLVSGR GTVGLNADNQ FNPDKLYGFG RGGRLDLANG SLSPHICNT
 601 TTKTNGRLNL VYQPAEEDRT LLLSGGTNIN GNITQNGKL FFSGRPTPHA
 651 YNHLNDHWSQ KEGIPRGETV WDNWDINRTF KAENFQIKGG QAVVSRNVAK
 701 VKGDWHLNSH AQAVFGVAPH QSHTICTRSD WTGLTNCVEK TITDDKVLAS
 751 LTKTDISGNV DLADHAHLNL TGLATLNGNL SANGDRTYTV SHNATQNGNL
 801 SLVGNAQATF NQATLNGNTS ASGNASFNLS DHAVQNGSLT LSGNAKANVS
 851 HSAALNGVSL ADKAVHFHES SRFTGQISGG KDTALHLKDS EWTLPSTGTEL
 901 GNLNLNATIT TINSAYRHDA AGAQTGSATD APRRRSRRSR RSLLSVTPPT
 951 SVESRFTLT NGKLNGQGT FRFMSELFGY RSDKLKLAES SEGTYTLAVN
 1001 NTGNEPASLE QLTVEGKDN KPLSENINFT LQNEHVDAGA WRYQLIRKDG
 1051 EFRLHNPVKE QELSDKLGA EAKKQAEKDN AQSLDALIAA GRDAVEKTES
 1101 VAEPARQAGG ENVGIMQAE EKKRVQADKO TALAKQREAE TRPATTAFPR
 1151 ARRARRDLPQ LQPPQPPQPP RDLISRYANS GLSEFSATLN SVFAVQDELD
 1201 RVFAEDRRNA VWTSGIRDTK HYRSQDFRAY RQCTDLRQIG MGKNLGSGRV
 1251 GILFESINTE NTFDDGIGNS ARLAHAVGG QGCLDFRIFG ISAGASFSGG
 1301 SLSDGIGGKI RRRVLYGKI ARYRAGFGCG GLEPHIGATR YFVQKADYR
 1351 ENVNIATPGL AFNRYRAGIK ADYSFKPAQH TSITPYLSLS YTDASGKVR
 1401 TRVNTAVLAQ DFGKTRSAEW GVNAEIKGFT LSLHAAAAGK PQLEAQHSAG
 1451 IKLGYRW*

Computer analysis of these sequences gave the following results:

25 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF1 shows 57.8% identity over a 1456aa overlap with an ORF (ORF1a) from strain A of *N.*

meningitidis:

30	orf1.pep	MKTTDKRRTTETHRKAPKGTGRIRFAXAAYLAICLSFGILLQAWAGHTYPGINYQYYRDAFEN	10	20	30	40	50	60
	orf1a							
	orf1a	MKTTDKRRTTETHRKAPKGTGRIRFSPAYLAICLSFGILLQAWAGHTYPGINYQYYRDAFEN	10	20	30	40	50	60
35	orf1.pep	KGKFVAGAKDIEVYNKKGELVGKSMTKAPMIDFSVVRNGVAALVGVQIVSVAHNGGYN	70	80	90	100	110	120
	orf1a							
	orf1a	KGKFVAGAKDIEVYNKKGELVGKSMTKAPMIDFSVVRNGVAALVGVQIVSVAHNGGYN	70	80	90	100	110	120
40	orf1.pep	NVDFGAEGXNIXDQXRXTYKIVKRNRYKAGTKGHYPGGDYHMPRLHKHVTDAEFVEMTSY	130	140	150	160	170	180
	orf1a							
	orf1a	NVDFGAEGXN-PDQHRFSYQIVKRNRYKPDNS-HEYNGDXHMPRLHKHVTDAEFVEMTSD	130	140	150	160	170	
45	orf1.pep	MDGRKYIDQNNYPDRVRIGRQYWRSDDEF-----NN-----	190	200	210			
	orf1a							
	orf1a	MRGNTYSDEKERYPERVRIGSGHHYWRYYDDKHGDLISYSGAWLIGGTHMQGWGNNGVXSL	180	190	200	210	220	230
50	orf1.pep	----RESSYH----IA-----SGSPMFITYDAQKQKWLINGVLQTGNPIYIGKSNQFQLVRK	220	230	240	250	260	
	orf1a							
	orf1a	SGDVRHANDYGPMPITAGAGDGSMPFIYDKTNNKWLINGVLQTGYGYPYSGRENGFQLIRK	240	250	260	270	280	290
55	orf1.pep	DWFYDEIFAGDTHSVFYEPQNGKYSFNDDNNGTGKINAKHEHSLNRLKTRTVQLFNW	270	280	290	300	310	320
	orf1a							
	orf1a	DWFYDDIYRGDTHTVXFEPRNSGHFSFTSNNNGTGTVTETNEKVSNP-KLKVQTVRLFDE	300	310	320	330	340	350
60			330	340	350	360	370	380

[illegible]

5	orf1.pep	-----					
	orf1a	LDLIAAGRDAAEKTESVAEPARKAGGENVIMGAAEEKKRVQADKDSALAKQREAE	TRP				
		1080	1090	1100	1110	1120	1130
						760	
10	orf1.pep	-----					
	orf1a	XTTAFPRARXARRDLQPQPQPQPQPQQRDLXSRYANSGLSEFSATLNSVFAVQDELD	R				
		1140	1150	1160	1170	1180	1190
15	orf1.pep	770	780	790	800	810	820
	orf1a	VFAEDRRNAVWTSQIRDTKHYRSQDFRAYRQOTDLRQIGMQKNLSSGRVGLFSHNRTEN					
		1200	1210	1220	1230	1240	1250
20	orf1.pep	830	840	850	860	870	880
	orf1a	TFDDGIGNSARLAHGA VFGQY GIDRFYIGISAGAGSSGSLSDGIGKXKRRVRLHYGIQA					
		1260	1270	1280	1290	1300	1310
25	orf1.pep	890	900	910	920	930	940
	orf1a	RYRAGFGGFGIEPIHIGATRYFVQKADYRYENVNIA TPGLAFNRYRAGIKADYSKPAQHI					
		1320	1330	1340	1350	1360	1370
30	orf1.pep	950	960	970	980	990	1000
	orf1a	SITPYLSLSYTDAA SGKVRTRVNTAVLAQDFGKTRSAEWGVNAEIKGFTLSLHAAA	K				
		1380	1390	1400	1410	1420	1430
35	orf1.pep	1010	1020				
	orf1a	QLEAQHSAGIKLYRWX					
		1440	1450				
40	orf1.pep	1010	1020				
	orf1a	QLEAQHSAGIKLYRWX					
		1440	1450				

The complete length ORF1a nucleotide sequence <SEQ ID 651> is:

45	1	ATGAAACAA	CCGACAAACG	GACAAACGAA	ACACACCCGA	AAGCCCCGAA	
	51	AACCGGCGC	ATCCGCTTCT	CGCGTGCTTA	CTTAGGCATA	TGCGTGTGCT	
	101	TCGGCATTCT	TCCCCAAGCT	TGGGCGGGAC	ACACTTATTT	CGGCATCAAC	
	151	TACCAATACT	ATCGCGACTT	TGCCGAAAAT	AAAGGCAAGT	TTGCAGTCGG	
	201	GGCGAAAGAT	ATTGAGGTNT	ACACAAAAAA	AGGGGAGATT	GTCCGCAACG	
50	251	CAATGACAAA	AGCCCGGATG	ATTGATTTT	CTCGGTGTC	CGCTAACGGC	
	301	CTGGCGGAT	TGTTGGCGGA	TCATATATT	GTGAGCGTGG	CACATAACGG	
	351	CGGCTATAC	AACGTTGATT	TTGGTGCGGA	AGGAAGNAAT	CCCGATCAGC	
	401	ACCGTTTTC	TTACCAAAAT	GTGAAAAGAA	ATAATATATA	GCCTGACRA	
	451	TCACACCTT	ACACACGGCA	TTANCAATAT	CGCGTTTTC	ATAAATTTGT	
	501	CACAGATGCA	GAACCTGTGG	AAATGACGAG	TGACATGAGG	GGGAATACCT	
55	551	ATTCCGATAA	AGAAAAATAT	CCCGAGCGTG	TCCGCTTCGG	CTCAGGACAC	
	601	CACATATTGC	GTTATGATGA	TGACAAACAC	GGCGATTAT	CCTACTCCGG	
	651	CGCATGGTTA	ATTGGCGGCA	ATACACATAT	GCAGGGTTGG	GGAAATAATG	
	701	GGGTANTTAG	TTTGAGCGGC	GATGTGCGCC	ATGCCAACGA	CTATGGCCCT	
60	751	ATGCCGATTG	CAGGTGCGGC	AGGCGACAGC	GGTTCGCCAA	TGTTTATTTA	
	801	TGACAAAACA	AACAATAAAT	GGCTGCTCAA	CGGAGTTTTA	CAAAACGGCT	
	851	ACCTTTATTC	CGGCAGGGAA	AACGGTTTCC	AGCTGATACG	CAAGATTTCG	
	901	TTCTACGATG	ACATTTACAG	AGGCGATACA	CATTACGCTC	NTTTTGACCG	
	951	CGCGATTAAC	GGACATTTAT	CCTTACATC	CTACGAGTAC	GTACGGGTTA	
65	1001	GTTAAACAGA	AACCAACGAA	AAGGTNTCCA	ATCCAAAGCT	TAAGATACAG	
	1051	ACAGTCCGAC	TGTTTGACGA	ATCTTTGAAT	GAACATGATA	AATCAACACT	
	1101	TTACCGCGCA	GGGGGTGTTA	ATCAGTACCG	TCGAAAGTTA	AACACCGGTG	
	1151	AAACACCTTC	TTTTATCGAT	TACGGCAACG	GCAAACTCAT	CTTATCAAAC	
	1201	AACATCAACC	AAGGCGCGGG	CGGTTTGAT	TTTGAAGGTG	ATTTACGGT	

1251 CTCGCCCTGAA AACAAACGAAA CGTGGCAAGG CGCGGGCGTT CATATCAGTG
 1301 AAGACAGTAG CGTTACATGG AATCTAAAGG CGGTGGCAAA CGACCGCGCT
 1351 TCCAAATACG GCAAAAGCAC GCTGCACGTT CAAGCCAAAG GGGAAACCTA
 1401 AGGCTCGATC AGCGTGGGCG ACGGTACAGT CATTTTGGAT CAGCAGGGAC
 1451 ACGATAAAGG CAAAAACAA GCCTTTAGTG AATTCGGCTT GNTCAGCGCG
 1501 AGGGGTACGG TGCACATGAA TGCCGATAAT CAGTTCACCC CCGACAAACT
 1551 CTATTTCGGC TTTCGGGCGG GACGTTTGA TTTAAACGGG CATTCGCTTT
 1601 CGTTCACCGG TATTCAAATAT ACGGATGAAG GGGCGATGAT TGNCNATCAT
 1651 AATGCCACAA CAACATCCAC CGTTACCAIT ACAGGGGAATG AAGATATTAC
 1701 ACAACCGAGT GGTAAAGATA TCAATAGACT TAATTACAGC AAAGAAATTG
 1751 CCTACAACGG TTGGTTTGGC GAGAAGATA CGACCAAAAC GAACGGCGCG
 1801 CTCACCTTGT TTTACGAGCC CGCGCGAGAA GACCGCACCC NGCTGCTTTC
 1851 CGCGGGRAACA AATTTAAACG GCAACATCAC GCAACCAACG GCAAACTGT
 1901 TTTTCAGCGG CAGACCGACA CGGACGCGCT ACAATCATTT AGGAAGCGGG
 1951 TGGTCAAAAA TGGAAGGTAT CCGACAAGGA GAATCGTGT GGGACAACGA
 2001 CTGGATCNAC CGCACGTTTA AAGCGGAAAA TTTCATATT CAGGGCGGGC
 2051 AGGCGGTGAT TTCCCGCAAT GTTGCCARAG TGAAGAGCGA TTGNCATTTG
 2101 AGCAATCAGC CCCAAGCAGT TTTTGTGCTC GCACCGCATC AAGCCATATC
 2151 NCTATTACGA CGATAAAGTG ATTGCTTCAT TGACTAAGAC NGACNTNAGC
 2201 GGCANTGTNA GNCNTNCCNA TNACGNTNNT TNAACNTCN CNNGGNTGCG
 2251 NNCACTNAAN GGCAATCTTA GTGCAAAATGG GATACACGTT TATACAGTCA
 2301 GCCACAACGC CACCCAAACG GGCACCTTTA GCCTCGTGGG CAATGCCCAA
 2351 GCAACATTTA ATCAAGCCAC ATTTAAACGGC AACNCACTCG NTTCCGGCAA
 2401 TGCTTCATTT AATCTAAGCA ACAACGCGCG ACRAAACGGC AGTCTGACGC
 2451 TTTCCGACAA CGCTAAGGCA ACGCTAAGCC ATTCCGACT CAACGGCAAT
 2501 GTCTCCCTAG CCGATAAGCG AGTATTCCAT TTTGAAACAA GCCGCTTTAC
 2551 CGGACAACCT AGCGGCAGCA AGGANACAGC ATTACACTTA AAAGACACGG
 2601 AATGGACGCT GCGCTCAGCG ACGGAATTAG GCAATTTAAA CTTTGACACG
 2651 GCCACCATTA CACTCAATTC CGCTATCGCG CACGATGCTG CAGGCGGGCA
 2701 AACCGGCAGN GTGTGACAGA CGCGCGCGCG CCGTTCGCGC CGTTCCCTAT
 2751 TATCGGTTAC ACCGCCAATC TCGGTAGAAAT CCGGTTTCGA CACGCTGACG
 2801 GTAAACGGCA AATTGAAGCG TCAGGAAGCA TTCGCTTTA TCTCGAATG
 2851 CTTCGGCTAC CGAAGCGGAA AATTGAAGCT GCGCGAAGT TCCGAGGNA
 2901 CTACACCTTT CATACCGCA ATTAACGGCA ACGAAACCGT AAGCTCTGAT
 2951 CAATTCACGG TAGTGCAGGG GAAAGACAC AAACCGCTGT CCGAAACCTC
 3001 TAATTTCACC CTGCAAAACG AACACGTGCA TGCGCGGCGG TGCGGTTACC
 3051 AACTCATCCG CAAAGACGGC GAGTTCGCGC TGCAATATCC GGTCAAAGAA
 3101 CAAGAGCTTT CGACAAACT CGGCAAGGCA GAAGCCAAAA AACAGCGGGA
 3151 AAAAGACAAAC GCGCAAGGCC TTGACGCGCT GATTGCGGCC GGGCGCGATG
 3201 CCGCCGAAAA GACAGAAAGC GTTGCGGAAC CGGCGCGGCG GGCAGCGGGG
 3251 GAAARTGTGC GCATTATGCA GCGGAGGAAA GAGAAAAAAC GGGTGCAGGC
 3301 GGATAAAGAC AGCGCNTTGG GAAACACGCG GAAGCGGAAA ACCCGCGCGG
 3351 NTACCACGCG CTTCGCCGCG GCCCGNCGG CCGCGCGGGA TTTGCGCCAA
 3401 CCGCAGCCCC AACCGCAACC TCAACCCCAA CCGCAGCGCG ACCTGATNAG
 3451 CCGTATGCCC AATAGCGGTT TGAGTGAATT TTCGCGCAGC CTCACACGCG
 3501 TTTTCGCGCT ACAGAGCGAA TTGACCGCGG TGTTTGGCGA AGACCGCGCG
 3551 AACGCGNTTT GGACAGCNGC CATCGGAGAC ACCTCAACTC ACGCTTCGCA
 3601 AGATTTCGCG CGCTACCGCG ACGGACCGCA CCGGGCCAAA ATTCGTTATG
 3651 AGAAACACTT CGGACGCGGG CGGCTGAGCA TCCTGTTTTC GCACAACGCG
 3701 ACCGAAACAA NCTTGCAGCA CGGCTCGCG AACTCGGCAC GGGTTGCCCA
 3751 CGGCGCGGTT TTCGGGCAAT ACGGCATCGG CAGGTTGCAC ATCGGATCA
 3801 GCACGGGCGC GGGTTTTCAG ACGGCGANTC TNCAGACGG CATCGGAGCG
 3851 AAAATCCGCG CGCGGCTGCT GCATTACGGC ATTCAGGCAC GATACCGGCG
 3901 CGGTTTCGCG GGAATTCGGA TCGAACCGTA CATCGGCGCA ACGCGCTATT
 3951 TCGTCCAAAA AGCGGATTAC CGCTACGAAA ACGTCAATAT CGCCACCCCC
 4001 GGTCTTCGCT TCAACCGNTA CCGNCGCGCG ATTAAGGCGAG ATTATTTCATT
 4051 CAACACCGCG CAACACATNT CCATCAACNC TTATTTNAGC CTGCTCTATA
 4101 CCGATCGCCG TCGGGGCAAA GTCCGAAAC CGGTCAATAC CGGNGTATTG
 4151 GCTCAGGATT TCGGCAAAAC CGGCAGTGCG GAATGGGGCG TAAACGCCGA
 4201 AATCAAAGGT TTCACGCTGT CCGTCCAGCG TGCGGCGCGC AAAGNGCGTG
 4251 AACTGGAAGC GCAACACAGC GCGGGCATCA AATTAGGCTA CGCGTGGTAA

This encodes a protein having amino acid sequence <SEQ ID 652>:

1 MKTIDKPTTE THRKAPKTR IRFSPAYLAI CLSFGILLPA WAGHTYFGIN
 51 YQYRRDFAEK GCKFAVGAKD IEDVNNKKEL VEGSMTKAEM IEDVSVSRNG
 101 VAALVGDDQYI VSAVHNGGVN NVDPAEGXNP PQHRSYSYQI VKRNNYKPSH
 151 SHPYNGDXHM PRLHKFVTDA EPEVMTSDNR GNTYSKERY PERVRIGSGDN
 201 HWRYDDDKH GDLSYSGAWL IGGNTHMQGW GNNGVXSLSG DVRHANDYGP

251 MPIAGAAGDS GSPMFIYDKT NNKWLNLGVL QTGYPSYSGRE NGFQLIRKDW
 301 FYDDIYRGDT HTVXFEPFRSN GHFSFTSNNN GTGTVTETNE KVSNGPKLRVQ
 351 TVRLFDDESIN ETDKEPVYAA GGVNQYRPRLL NNGENLSFID YGNGKLILSN
 401 NINQGAGGLY FEGDFTVSPFE NNETWQAGAV HISEDSTVW KVNQVANORL
 451 SKIKGTLLHV QAKCEQGGSI SVGGTGVLLD QQADSGKKQ APSBGLKSG
 501 RGTQVLNADN QPNEDKLYFG FNGSRLLDNG HSLSFHRIQN TDEGAMIXH
 551 NATTTSTVTI TGNESITQPS GKNINRLNYS KEIAYNGWEG EKD'TKTNGR
 601 LNLVYQPAAE DRTXILSGGT NLNGNITQTN KGLFFSGRPT PHAYNHLGSG
 651 WSKMEGI PQG EIVVNDNDWIX RTFKAENFHI QGGQAVISRN VAKVEGDGXL
 701 SNHAQAVFGV APOHSHTTCT RSDWTGLTNC VEXXI'DDKV IASLTKTDXS
 751 GXVKLXXXXX XXLGXGAXLK GNLGANGDTR YTVSHNATQN GNLVLGNAQ
 801 ATPNQATLNG NKXSXGNASF NLSNNAQNG SLTSLDNAKA NVSHSALNG
 851 VSLADKAVFH FENSRTGQL SGSKXTALHL KDSEWTLPSG TELGNLNLDN
 901 ATITLNSAYR HDAAGAQTGX VSDTPRRRS RSLLSVTPPT SVESRFTLTL
 951 VNGKLNXQST FRFMSLFGY RSDKLKLAES SEGTYTLAVN NIGNEPVSLD
 1001 QLTVVEGKDN KPLSENLFNT LQNEHVDAGA WRVQLIRKDG EFRLHPVKE
 1051 QELSDKLGA EAKKQAEKDN AQSLDALIAA GRDAEKTES VAEFARXAGG
 1101 ENVGIMQAE EKKRVQADKD SALAKQREAE TRPKXTAFPR ARXARRDLQ
 1151 PQPQPQPQPQ PQRDLSXRYA NSGLSEFSAT LMSVFAVQDE LDRVFAEDRR
 1201 NAVVTSEKTR TKHYRSDER AYRQQTDLRG IGMQNLGSG RVGILFSHNR
 1251 TENKFDGGS NSAKLHSAV FGQYGGHGF IQISTGAGFS SGXLSGJGG
 1301 KIRRRVLHYG IQARYRAGFG GEGIEPYTGA THYFVQKADY RVENVNATP
 1351 GLAFNRYRAG IKADYSFKPA QHXSTPYXS LSY'DAASGK VTRVNTAVL
 1401 AQDPGKTRSA EWGVNAEIKG FTLXHXAAAA KGPQLEAQS AGIRLGYRW*

A transmembrane region is underlined.

ORF1-1 shows 86.3% identity over a 1462aa overlap with ORF1a:

		10	20	30	40	50	60
orfla.pep		MKTTDKR	TTETHRK	APKTGR	IRFSPAY	LAICLSF	GILPQAWAGHTYFGINQYYRDF
orfl-1		MKTTDKR	TTETHRK	APKTGR	IRFSPAY	LAICLSF	GILPQAWAGHTYFGINQYYRDF
		70	80	90	100	110	120
orfla.pep		KGKFAVG	AKADIEV	NKKGEL	VGKSMTK	KAMIDF	SVVSRNGVAALVGDQYIVSVAHNGGYN
orfl-1		KGKFAVG	AKADIEV	NKKGEL	VGKSMTK	KAMIDF	SVVSRNGVAALVGDQYIVSVAHNGGYN
		130	140	150	160	170	179
orfla.pep		NVDFGAE	GXNPDQ	HRFSYQ	IVKRNNY	KPDNS-HPYNGDXHMPRLHKFVTD	AEPEVMTSDM
orfl-1		NVDFGAE	GRNPDQ	HRFTYK	IVKRNNY	KAGTKGHFYGGDYHMPRLHKFVTD	AEPEVMTSYM
		180	190	200	210	220	230
orfla.pep		RGNTYS	DKKEKY	EEVRVY	IGSGHHY	WRYYDDKHGDL--SYSGA----	WLI GGNTHMQGWGNN
orfl-1		DGRKYI	QNNYPD	RVRIAG	RGYWR	SDDEFPNNRESSYHIASAYSWLVGGNT	FANQSGGG
		240	250	260	270	280	290
orfla.pep		GVXLSGD	-VRHANDY	GPMP	PIAGAAG	DGSGSPMFIYDKTNNKWLNLGVL	QTGYPSYSGRENG
orfl-1		GTNVLG	SEKIKHS-PYGFLT	CGSGFG	DGSGSPMFIYDAQKQKWLNLGVL	QTGNFYPIKSGN	
		300	310	320	330	340	350
orfla.pep		FOLIRKDW	FYDDIYRG	DTHTVX	FEPRSN	GHFSFTSNNGTGTVTETNEKVSNNP	-KLKVQT
orfl-1		FQLVKDW	FYDEIFAG	DTHSEF	EPQNGK	SYNDDNNGTGTGTINAKHEHNSL	PNRLKTRT
		360	370	380	390	400	410
orfla.pep		VRLFDES	INETDKEP	VY-AAGG	VNQYRPR	LLNNGENLSFIDY	GNGLKLLSNINQGAGGLY
orfl-1		VQLFNVS	LSETAREP	VYHAAG	VNSYRPR	LLNNGENISFIDEGK	GELILT'SNINQGAGGLY

	360	370	380	390	400	410
5		420	430	440	450	460
orf1a.pep		FEQDFTVSPENN	ETWQAGVHISED	STVTWKVNGVAND	RLSKIGKGT	LHVQAKGENCG
orf1-1		FEQDFTVSPENN	ETWQAGVHISED	STVTWKVNGVAND	RLSKIGKGT	LHVQAKGENCG
10		480	490	500	510	520
orf1a.pep		SVGDGT	VILDQQADDK	GKQAFSEI	GLXSGRGT	VQLNADNQFN
orf1-1		SVGDGT	VILDQQADDK	GKQAFSEI	GLXSGRGT	VQLNADNQFN
15		540	550	560	570	580
orf1a.pep		HSLSFHRIQNT	DEGAMIXXHN	ATTSTVTIT	CNESITQ	PSGKNNIR
orf1-1		HSLSFHRIQNT	DEGAMIXXHN	ATTSTVTIT	CNESITQ	PSGKNNIR
20		600	610	620	630	640
orf1a.pep		EKDTTKTNGR	LNLVYQPAE	DRTXLLSGG	TNLTNGIT	QTNGKLF
orf1-1		EKDTTKTNGR	LNLVYQPAE	DRTXLLSGG	TNLTNGIT	QTNGKLF
25		660	670	680	690	700
orf1a.pep		WSKMEGIPQGE	IWVNDWIXR	TFKAENFHI	QGGQAVIS	RNVAKV
orf1-1		WSKMEGIPQGE	IWVNDWIXR	TFKAENFHI	QGGQAVIS	RNVAKV
30		720	730	740	750	760
orf1a.pep		APHQSHTICTR	SDWTGLTNC	VEVXITDD	KVVIASLT	KTDXSGK
orf1-1		APHQSHTICTR	SDWTGLTNC	VEVXITDD	KVVIASLT	KTDXSGK
35		780	790	800	810	820
orf1a.pep		GNLSANGDTRY	TVSHNATQ	NGNLSLVG	NAQATFNQ	ATLNGNX
orf1-1		GNLSANGDTRY	TVSHNATQ	NGNLSLVG	NAQATFNQ	ATLNGNX
40		840	850	860	870	880
orf1a.pep		SLTSLDNK	ANVSHS	ALNGVSL	ADKAVH	EFENSR
orf1-1		SLTSLDNK	ANVSHS	ALNGVSL	ADKAVH	EFENSR
45		900	910	920	930	940
orf1a.pep		TELGNLNL	NATITL	NSAYRH	DAAGTQ	GXVSDT
orf1-1		TELGNLNL	NATITL	NSAYRH	DAAGTQ	GXVSDT
50		950	960	970	980	990
orf1a.pep		TLTVNGKL	NKQGT	FRFMS	ELFGYR	SDKCLK
orf1-1		TLTVNGKL	NKQGT	FRFMS	ELFGYR	SDKCLK
55		1010	1020	1030	1040	1050
orf1a.pep		KDNKPL	SENIN	FTLQ	NHVD	AGAWRY
orf1-1		KDNKPL	SENIN	FTLQ	NHVD	AGAWRY
60		1070	1080	1090	1100	1110
orf1a.pep		KDNAQSL	DALIA	AGRDAA	EKTES	VAEPA
orf1-1		KDNAQSL	DALIA	AGRDAA	EKTES	VAEPA

[illegible]

Homology with adhesion and penetration protein hap precursor of *H.influenzae* (accession number P45387)

Amino acids 23-423 of ORF1 show 59% aa identity with hap protein in 450aa overlap:

orf1	23	FXXAAVLELISFGLIPQAWAGHTYFGFVINYQYRDFAEKNKGKFFVAGAKDIEVNVKKGLEVG	82
hap	6	F + L C4S GI QAWAGHTYFGF+YQYRDFAEKNKGKFF VGAK+IEVNVK+G+LVG	
		FRILNFLTACVSLGSIASQAWAGHTYFGDIDYQYRDFAEKNKGKFFVGAKNIEVNVKEGQVLG	65
orf1	83	KSMTKAPMIDFVSVRNRGVAALVGQYIVSVAHNGGNNYNDVFGAEGXNIDQXRTYKIV	142
hap	66	SMTKAPMIDFVSVRNRGVAALVG QYIVSVAHNGGNN+VDFGAEG N D Q R X R T Y+V	
		TSMTKAPMIDFVSVRNRGVAALVGQDIIVSVAHNGGNNDFVFGAEGRN+PDHQHRTYQV	124
orf1	123	KRNNYKAGTKGHFGYGGDYHMLPRHLKXVDEAPEVMTSYMDGRKYIDQNNYPRVRVIGAG	204
hap	145	KRNNY+A + HPY GDYHMLPRHLK VT+AEVP MT+ MDG+ Y D+ NYP+RVRI+GR	
		KRNNYQAWERKHFGYGGDYHMLPRHLKHFVTEAEPVGMTTNMDGKVYADENYPERVIRIGSR	184
orf1	203	QYWRSDDEDFNNRESSYHIA	222
hap	185	QYWR+D+DE N SSY	
		QYWRDSDDEETNVSYSYVSGAYRYLTAGNTHTQSGNGNGTVNLSGNVVSFNHYGLPTG	244
orf1	223	-----SGSPMFITYDAQKQKLLINGVLQGNPFYIGKSGNQLFVRKDDVFDIEFAGHTSV	277
hap	245	SGSPMFITYDA+K+HLLIN VLQTG+P+ G NQFL+R+R+WFY+E+ A DT SVF	
		SGSGDQSGSPMFITYDAKKDKLLINAVLQTGHGHPFFGRNGNQLIREWFYFNEVLAVSTFVS	304
orf1	278	---YEPNRQKYSFDDNCTGCKLIN-AKHEHNSLNRKLTIRTVNLSLSTATSPVPYHA	344
hap	305	Y P NG YSF +N+GTGK+ + TVL+LNSL+T+T+E+ V A	
		ORXYPEINGHYSFDDNCTGKILITRFSKDGSKASFEVGTVLNLSLSTATAKHY+KA	363

orf1 335 AGGVNSYRPRLNNGENISFIDEGKGLILTSNINQAGGLYFGQDFTV-SPENNETWQGA 393
 A G N Y+P+ G+NI D+GKG L + +NINQAGGLYF+G+V F +NN TWQGA
 hap 364 AAGYNYIQPRMEYGRNIYLDGQKGTLTIENNINQAGGLYFEGNEFVVKQNNITWQGA 423

5 orf1 394 GVHISEDSTVTWKVNGVANDRLSKIGKGT L 423
 GV I +D+TV WKV+ NDRLSKIG GTL
 hap 424 GVSIGQDATVEWKVHNPENDRLSKIGIGTL 453

Amino acids 715-1011 of ORF1 show 50% aa identity with hap protein in 258aa overlap:

Orf1 41 DTRYTVSHNATQ-NGNXSLVXNAQATFNQ-ATLNGTASAGNASFNLSDHVAQNGSLTLS 98
 DT+ S TQ NG+ +L NA + A LNGN + ++ F LS++A Q G++ LS
 hap 733 DTKVINSIPITQINGSINLTNNATVNIHGLAKLNGVTLIDHSQFTLSNNATQTGNIKLS 792

orf1 99 GNAKANVSHSALNGVSLADKAVPHFESSRFTGQISGGKDTALHKLKDESWTLPSGXELGN 158
 +A A V+++ LNGNV L D A F ++S F QI G KDT + L+++ WT+ES L N
 hap 793 NHANATVNNATINGNVHLTDSAQFSLKNSHFHWQIQGDKDOTTVTLENATWTFSDTTIQN 852

orf1 159 LNLONATITLNSAYRHDAGAQTGSATDAPKXXXXXXXXXXLLXVTPPTSVESRENTLTVN 218
 L L+N+T+ZLSAY + S+ +P L T PTS E RENTLTVN
 hap 853 LTIANNSTVTLNSAY-----SASSNNAPRHRS-----LETETTTTSAEHRFNTLTVN 899

orf1 219 GKINGQOTFRFMSELFGRSDKLKLAESSEGTVT LAVNNTGNEPASLEQLTVVEGDKNPK 278
 GKL+GQQTf+F S LFGY+SDKLKl+ +EG YTL+V NTG EP +LEQLT++E DNPK
 hap 900 GKLSGQGTfQFTSLSLFGYKSDKLKLSNDAEGDYTLVSRNTGKEPVTLEQLTLIESLDNPK 959

orf1 279 LSENINFTLQNEHVDAGA 296
 LS+ L FTL+N+HVDAGA
 hap 960 LSDKLKFTLENDHVDAGA 977

Amino acids 1192-1450 of ORF1 show 41% aa identity with hap protein in 259aa overlap:

Orf1 1 LDRVFAEDRRNAVVTSGIRDTKHYRSQDFRAYRQOTDLRQIGMQKNLGSRGVILFHSNR 60
 LDR+ F + ++AVWT+ +D+ Y S FRAY+Q+T+LRQIG+QK L +GR+G +FSH+R
 hap 1135 LDRLFVDQQAQSAVWNTINIAQDKRYSDBAFRAYQQTNLRLQIGVQKALANGRIGAVFHSR 1194

orf1 61 TENTFDDGIGNSARLAHGAVFQYQIGIDRFYXXXXXXXXXXXXXXXXXIGKXRRRVLHYG 120
 ++NTFD+ + N A L + F QY K R+ ++YG
 hap 1195 SDNTFDEQVKNHATLTMSGFAQYQWGDLPQGVNVGTGISASKMAEEQSKRIHRKAINYG 1254

orf1 121 IQARYRAGFGGFGIEPHIGATRYFVQKADRYENVNIAITPGLAFNRYRAGIKADYSFKPA 180
 +A Y+ G GI+P+ G RYF+++ +Y+ E V + TP LAFNRY AGI+ DY+F P
 hap 1255 VNASYQFRLQQLGIQYFPGVNRVFIERNYQSEEVVRKTPSLAFNRYNAGIRVDYTFPT 1314

orf1 181 QHITSITPYSLSYTDAAAGKVRTRVNTAVLAQDFGKTRSAEWGVNAIEKGTFTSLHAAAA 240
 +IS+ PY ++Y D ++ V+T VN VL Q FG+ E G+ AEI F +S + +
 hap 1315 DNISVKPFFVNVYVDVSNANVTQTNVLTVLQFPFGRYWQKEVGLKAEILHFQISAFISKS 1374

orf1 241 KGPQLEAQSAGIKLGYRW 259
 +G QL Q + G+KLGYRW
 hap 1375 QGSQGLKQQNVGVKLGIRW 1393

Homology with a predicted ORF from *N.gonorrhoeae*

The blocks of ORF1 show 83.5%, 88.3%, and 97.7% identities in 467, 298, and 259 aa overlap, respectively with a predicted ORF (ORF1ng) from *N.gonorrhoeae*:

orf1.pep MKTTDKRTTETHRKAPKTGRIRFXAAYLAICLSFGLPQAWAGHTYFGYNGIYYRDFASN 60
 orf1ng MKTTDKRTTETHRKAPKTGRIRFSPAYLAICLSFGLPQARAGHTYFGYNGIYYRDFASN 60

orf1.pep KGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFVSVSRNGVAALVGQYIVSVAHNGGYN 120
 orf1ng KGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFVSVSRNGVAALAGDQYIVSVAHNGGYN 120

orf1.pep NVDFGAEGNIXDQXRTYKIVKRNNYKAGTNGHPYGGDYHMPRLHKVFDAEPEVENTSY 180
 orf1ng NVDFGAEGSN-PDQHRFSYQIVKRNNYKAGTNGHPYGGDYHMPRLHKVFDAEPEVENTSY 179

5	orf1. pep	MDGRKYIQNNYPDRVRIGAGRQYWRSDEDEFPNNRESSYHIAS-----	223
	orf1ng	MDGWKYADLNKYPDRVRIGAGRQYWRSDEDEFPNNRESSYHIASAYSWLVGGNTFAQNGSG	239
10	orf1. pep	-----GSPMFIYDAQKQKWLINGVLQTGNPYIGKSNG	255
	orf1ng	GGTVNLGSEKIKHSPYGLPTGGSGFGDSGSPMFIYDAQKQKWLINGVLQTGNPYIGKSNG	289
15	orf1. pep	FQLVRKDWFYDEIFAGDTHSVFYEPHQNGKYSFNDNNNGTGKINAKHEHNSLPNRLKTRT	315
	orf1ng	FQLVRKDWFYDEIFAGDTHSVFYEPHQNGKYFFNDNNNGAGKIDAKHKHYSLPYRLKTRT	359
20	orf1. pep	VQLFNVSLSSETAREPVYHAAGGVNSYRPRLNNGENISFIDEGKGELILTSNINQAGAGGLY	375
	orf1ng	VQLFNVSLSSETAREPVYHAAGGVNSYRPRLNNGENISFIDEGKGELILTSNINQAGAGGLY	422
25	orf1. pep	FQGDFTVSPENNETWQAGVHI SEDSTVTWKVGVANDRLSKIGKGT	442
	orf1ng	FEGNFTVSPKNNETWQAGVHI SDGSTVTWKVGVANDRLSKIGKGLTLVQAKENGQGSV	479
30	orf1. pep	// DKVTASLTKTDISGNVDLADHAHNLITGLA	744
	orf1ng	FGVAPHQSHITICRSDWTGLTSCTEKTI DDKVIAISLKT DVRGNVSLADHAHNLITGLA	774
35	orf1. pep	TLNGNLSANGDTR--YTVSHNATQNGNXSLVXNAQAATFNQATLNGNTSASGNASFNLSDHA	803
	orf1ng	TFNGNL-VQAETRTIRLRANATQNGNLSLVGNAQAATFNQATLNGNTSASGNASFNLSNNA	833
40	orf1. pep	VQNGSLTSGNAKANVSHSALNGVNSLADKAVHFHFSRRTGQISGGKDTALHLKDSWT	863
	orf1ng	VQNGSLTSDNAKANVSHSALNGVNSLADKAVHFHFSRRTGKISGGKDTALHLKDSWT	893
45	orf1. pep	LPSGKELGNLNDNATITLNSAYRHDAAAGATGGSATDAPRRSRRRSRRLSXVTPPTSVE	923
	orf1ng	LPSGKTELGNLNDNATITLNSAYRHDAAAGATGSAADAPRRSRRRS---LSVTPPTSVE	950
50	orf1. pep	SRNFNTLVNGKLNQGTFRFMSELFGYRSDKLLKLAESSEGTYTLAVNNTGNEPASLEQLT	983
	orf1ng	SRNFNTLVNGKLNQGTFRFMSELFGYRSGKLLKLAESSEGTYTLAVNNTGNEPVSLEQLT	1010
55	orf1. pep	VVEGKDNKPLSENLFNTLQNEHVDAGAW	1011
	orf1ng	VVEGKDNTPLENLNFNTLQNEHVDAGAWRYQLIRKDGFEFLNHPVKEQELSDKLGAGET	1070
60	orf1. pep	// LDRVFAEDRRNAVVTSGIRDTKHYRSQDFR	1211
	orf1ng	PQRDLISRYANSGLSEFSATLNSVFAVQDELDRVFAEDRRNAVVTSGIRDTKHYRSQDFR	1239
65	orf1. pep	AYRQQTDLRQIGMQKNLGSGRVGLFHSNRTENTFDDGIGNSARLAHGA VEGGYGIDREY	1271
	orf1ng	AYRQQTDLRQIGMQKNLGSGRVGLFHSNRTGNTFDDGIGNSARLAHGA VEGGYGIGRFD	1299
70	orf1. pep	IGISAGAGFSSGSLSDGIGKXRRRVLHYGIQARYRAGFGGFGIEPHIGATRYFVQKADY	1331
	orf1ng	IGISAGAGFSSGSLSDGIRGKIRRRVLHYGIQARYRAGFGGFGIEPHIGATRYFVQKADY	1359
75	orf1. pep	RYENVNIATPGLAFNRYRAGIKADYSFKPAQHISITPYLSLSYTDAAAGKVRTRVNTAVL	1391
	orf1ng	RYENVNIATPGLAFNRYRAGIKADYSFKPAQHISITPYLSLSYTDAAAGKVRTRVNTAVL	1419
80	orf1. pep	AQDFGKTRSAEWGVNAEIKGFTLSLHAAAAGKPOLEAQHSAGIKLGRYW 1440	
	orf1ng	AQDFGKTRSAEWGVNAEIKGFTLSLHAAAAGKPOLEAQHSAGIKLGRYW 1468	

The complete length ORF1ng nucleotide sequence was identified <SEQ ID 653>:

```

1  ATGAAACAA  CCGACAAACG  GACAACCGAA  ACACACCGCA  AAGCCCCCTAA
51  AACCCGCCGC  ATCCGCTTCT  CGCCCGCTTA  CTTAGCCATA  TGCCTGTGCT
101 TCGGCATCTT  GCCCAAGCC  CGGGCGGGAC  ACACCTATT  CGGCATCAAC

```

151 TACCAATACT ATCGCGACT TCGCGAAAT AAAGGCAAGT TTGCAGTCGG
 201 GGCGAAAGT ATTGAGGTTT ACAACAAAA AGGGGAGTTG GTGCGCAAT
 251 CGATGACGAA AGCCCCGATG ATTGATTTTT CTGTGGTATC GTGTAACGGC
 301 GTGGCGGCAT TGGCGGGCGA TCAATATATT GTGAGCGTGG CACATACCGG
 351 CGGCTATAAC AATGTTGATT TTGGTGCGGG GGGAGCAAT CCCGATCAGC
 401 ACCGCTTTTC TTACCAAAAT GTGAAAGAA ATAAATTATA AGCAGGGACT
 451 AACGGCCATC CTTATTGGCG CGATTATCAT ATGCGCGT TTGCACAAAT
 501 TGTACAGATG GCAGAACCTG TTGAGATGAC CAGTTATATG ATGGGTGGA
 551 AATACGCTGA TTTAAATAAA TACCCTGATC GTGTCTGAAT CGGAGCAGGC
 601 AGACAATATT GGCGGTCTGA TGAAGACGAA CCCAATAACC GCGAAAGTTC
 651 ATATCATATT GCAAGCGCAT ATTCTTGCT GTCTGGTGGC AATACCTTTG
 701 CACAAATAGG ATCAGGTGGT GGCACAGTCA ACTTAGGTAG CGAAAAATTC
 751 AAACATAGCC CATATGGTTT TTTACCAACA GGAGGCTCAT TTGGCGACAG
 801 TGGCTACCCA ATGTTTATCT ATGATGCCCA AAAGCAAAAG TGGTTAATTA
 851 ATGGGGTATT GCAACAGCGC AACCCCTATA TAGSAAAAAG CAATGSCCTC
 901 CAGCTAGTTC GTAAAGATTG GTTCTATGAT GAAATCTTTG CTGGAGATAC
 951 CCATTACGTA TTCACGCGAC CACATCAAAA TGGGAATAC TTTTITACGC
 1001 ACAATAATAA TGGCGCAGGA AAAATCGATG CCAACMTAA ACATATTTCT
 1051 CTACTCTATA GATTAAANAC ACGAACCGTT CAATTGGTTA ATGTTCTCTT
 1101 ATCCGAGACA CGAAGAGAAC CTGTTTATCA TGCTGCAGGT GGGGTCAACA
 1151 GTTATCGACC CAGACTGAAT AATGGAGAAA ATATTTCTTT TATTGCACAA
 1201 GGAANAAGTG AATTGATACT TACCAGCAAC ATCAACCAAG GCGCGGGCGG
 1251 TTTGTATTTT GAGGGTAAAT TTACGGTCTC GCCTAAAAAC AACGAAAGCT
 1301 GGCAAGGCGC GGGCGTTTAT ATCAGTGATG CGAGTACCGT TACTTGGAAA
 1351 GTAAAGCGCG TGGCAACGGA CCGCCTGTCC AAAATCGGCA AAGGCACGCT
 1401 GCTGGTTCAA GCCAAAGGGG AAAACCAAGG CTCGGTCAGC GTGGGCGACG
 1451 GTAAAGTCAT CTTAGATCAG CAGCGCGAGC ATCAAGSCAA AAAACAGCGC
 1501 TTTAGTGAAA TCGGCTTGGT CAGCGCGAGG GAGACGGTGC AACTGAATGC
 1551 CGATAATCAG TTCAACCCCG ACAACCTCTA TTTGCGGTTT CGCGCGGACG
 1601 GTTTGGATTG GAACGGGCAT TCGCTTTGCT CAACCGCAT TCAAAATACC
 1651 GATGAAGGGG CGATGATTGT CAACCAACAT CAGACCAAG ARTCCACGCT
 1701 TACCATATCA GGCATAAAG ATATTCTATC AACCGCAAT AACAACGACT
 1751 TGGATAGCAA AAAAGAAAT GCCTACACG GTTGGTTTGG CGAGAAAGAT
 1801 GCAACAGGCA CGAACGGGAT CTCTCAATCT AAATTACCAAC AGAGGAAGCG
 1851 TGATCGCACT TTACTCGTCT CCGCGGAAC AATTTTAAGG CGCATATACA
 1901 CGCAAAACAA CGGCAAACTG TTTTCCAGCG GCAGACCGAC ACCGCGACCG
 1951 TACAATCATT TAGSAAAGCG GTGGTCAAAA ATGGAAGGTA TCCCAACAAG
 2001 AGAAATCTGT TGGGACAAGC ATTGGATCGA CGGCAATTT AAAGCGGAAA
 2051 ACTTCCATAT TCAGGGCGGA CAAGCGGTGG TTTCCGCAAA TGTGTCACAA
 2101 GTGGAAGGCG ATTGCACTTT AAGCAATCAC GCCCAAGCAG TTTTCGGTGT
 2151 CGCACCGCAT CAAAGCCACA CAATCTGTAC ACGTTGGGAC TGGACGGGTC
 2201 TGACAAGTTG TACCGAAAA ACCATTACCG ACGATAAAGT GATTGGCTCA
 2251 TTGAGCAAGA CGACATCAG AGGCAATGTC AGCCTTGCCG ATCACGCTCA
 2301 TTTAAATCTC ACAGGACTTG CCACACTCAA CGGCAATCT AGTGACGGCT
 2351 GAGACACGCA CTATACGGTT CGGCGCAAGC CCAACAAAA CGGCAACTTC
 2401 AGCCTCGTGG GCATAGCCCA AGCAACATTT AATCAAGCCA CTATCAAGCT
 2451 CACACATGCG GCTTCGGACA ATGCTTCATT TAATCTAAGC AACRACGGCG
 2501 TACAAACGCG CAGTCTGACG CTTTCCGACA AGCCTAAGCG AAAGCTAAGC
 2551 CATTCCGCACT CAACGGCACA TGCTCCCTTA GCGGATAAGG CAGTATTCCA
 2601 TTTTGAARAC AGCCGCTTTA CCGSAAAAAT CAGCGCGCGC AAGGATACGG
 2651 CATTAACACT AAAAGACAGC GAATGGAAGC TGCGTCCGGC CACGGAATTA
 2701 GGCAATTTAA ACCTTGACAA CGCCACCAAT ACACTCAATT CGCCTATCG
 2751 ACACGATGCG GCAGGCGCGC AAACCGGCAAG TGCGGAGAT CGCGCGCGCG
 2801 GCGGTTTCGG CCGTTCCTTA TTATPCGTTA CGCGCGCAAC TTGCGCAGAA
 2851 TCCCGTTTCA ACACGCTGAC GGTAAACGGC AAATTTGAAC GTACGGGAGC
 2901 ATTCTCGCTT ATGTGGAAC TCTTCGGCTA CCGCAGCGCG AAATTTGAAG
 2951 TGGCGGAAAG TTTCGAAGGC ACTTACACCT TGGCTGTCAA CAATACCGCG
 3001 AACGAACCGG TAAGTCTCGA GCAATTGACG GTAGTGAAG GAAAAGACAA
 3051 CACACCGCTG TCGGAAATCT TTAATTTTAC CCGTGCaaAc gaacacgtgcy
 3101 atgcgcggcgc atggCGTTAT CAGCTTATCC gcaasagcgc CGAGTCTCgc
 3151 CTGCATAATC CGGTCAAAGA ACAGAGACTT TC CGCAAAAC TGCGCAAGc
 3201 gggagaaACA gggcgcgct TGA CGCGCAA ACAGCGcaAA CTTCGCGCA
 3251 AACacacggc gjaaaAAGAC AACcgcgaaa gcttgcgcgc gctgattgcg
 3301 gCcgggcgca atgccaccga AAAGSGAga aqgttgcgc aaecgcGCCG
 3351 CGAGGCGAGC GGGGAAATg ccgGCA'TAT CAGCGCGGAG GAAGGAJAAA
 3401 AACGGGTGCA GCGGATAAAA GACACCGCCT TGGCGAAACA CGCGGAAGCG
 3451 GAAACCGGCG CGGCTACCAAC CGCCTTCCCG CGCGCCCGCG CGCGCGAGCG
 3501 GGATTTGCGC CAACCGCAGC CCGAACCGCA ACCCGAACCG CAGCGCGACC
 3551 TGATCAGCGG TTATGCCAAT AGCGGTTTGA GTGAATTTTC CGCCACGCTC
 3601 AACAGCGTTT TCGCCGTACA GGACGAATTG GACCGCGTGT TTGCGGAGAA

CHIR-0160 (356.001) PATENT

3651 CCGCCGCAAC GCGGTTTGA CAAGCGGCAT CCGGGACACC AAACACTACC
 3701 GTTCGCAAGA TTTCGGGCC TACCGCCAAC AAACCGACCT GCGCCAAATC
 3751 GGTATCGAGA AAAACCTCGG CAGCGGGCCG GTCCGGCATC TGTTCGCAAC
 3801 CACCGCGACC GGAACACCT TCACCGACAG CATCGGCATC TCGCGCACAT
 3851 TTGCCCGAGG TGCGGTTTTC GGGCAATACG GCATCGGCAG GTTCGACATC
 3901 GGCATCAGCG CCGGCGCGGG TTTTAGTAGC GGCAGCCTTT CAGACCGCAT
 3951 CAGAGGCAAA ATCCCGCGCC GCGTGTCTCA TTACGCATTT CAGGCAAGAT
 4001 ACCGCGCAGG TTTCGGCGGA TTCCGTCATC AACCGCACAT CCGCGCAACG
 4051 CGCTATTTCG TCCAAAAGC GGATTACCGA TACGAAACG TCAATATCGC
 4101 CACCCGGGCG CTTCGATTCA ACGCTACCG CCGGGCGATT AAGGCAGATT
 4151 ATTCATTCAA ACGCGCGCAA CACATTCCA TCAGCGCTTA TTTGAGCTCG
 4201 TCCTATACCG ATCGCGCTTC CCGCAAAGTC CGAACGCGCG TCAATACCGC
 4251 CGTATTGGCG CAGGATTTCG GCAAAACCCG CAGTCGCGAA TGGGCGCTAA
 4301 ACGCCGAAT CAAGGTTTC ACGCTGCC TCACGCTGC CCGCGCCAGC
 4351 GGGCCGAAT TGGAAGCGCA GCAAGCGCG GGCATCAAT TAGGCTACCG
 4401 CTGTTAA

This is predicted to encode a protein having amino acid sequence <SEQ ID 654>:

1 MKTTDKRTTE THRKA¹PKTGR IRFSPAYLAI CLSPGILPQA RAGHTYFGIN
 51 YQYRDPFEN KKGPAVGAKD IEVYNKGGEL VGGSMTKAPM IDFSVVSRRNG
 101 VAALAGDQYI VSAHNGGYN NVDFGAE²GSN PQHRSFYGI VKRNNYKAGT
 151 NGHPYGGDYH MPRLHKFTVD AEPVEMTSYM DQWKYADLNK YFVRDRIAG
 201 RQYWRSDDEE PNNRESSYHI ASAYS³WLVGG NTFANQSGGG GTVNLGSEKI
 251 KHSPLYGLEP GSGFGDSGSP MFIYDAQKQK WLINGVLQTG NFIYIGSKNGF
 301 QLVRKDWDFY EIFAGDTHSV FYEPHONGKY FNDNNNGAG KIDAKHKHYS
 351 LPYRLKTRTV QLFNVLSET AREPVYHAAG GVNSYRPLRN NGENISFTDK
 401 KGKGLILTSN INQAGGGLYF EGNFTVSPKN NETWQAGVH ISDGSFTWVK
 451 VNGVANDRLS KIGKGLTLVQ AKGENQGSVS VGGDKVILDQ QADDQCKKQA
 501 PSEIGLVSGR GTVQLNADNQ FNPDKLYPGF RGGRLDLNGH SLSPHRIOT
 551 DEGAMIVNHN QDKESTVTIT GNKDITTTGN NNNDLSKKEI AYNGWFGSKD
 601 ATKNTNGLNL NYPPERADRT LLLSGSTNLN GNITQTNKGL FFSGRPTPHA
 651 YNHLSSGWSK MEGIPQGEIV WDNDWIDRTF KAENFHIQGG QAVVS⁴RNVAK
 701 VEGWHLNSH AQAVFGVAPH QSHTICTRSD WGLTSCTEK TITDDKVTAS
 751 LSKTIDVRSN SLADRAHML TGLATFENGL VQAE⁵TRTIRL RANATQNGNL
 801 SLVGNQATFT NQMTLNGMVS ASDNASFNLS NNNAVQNGSL LSDNAKANVS
 851 HSA⁶LN⁷GNVSL ADKAVPHEN SRPTGKISGG KDTALLHLKDS EWTLPSCGLE
 901 GNILNDLNATI TLNSAVRHDA AGAQTC⁸SAAD APRRRSRSL LSVTPPTSAB
 951 SRFNLTIVNG KLNQGTFRF MSELFCYRSG KLKLAESSEG TYTLAVNWTG
 1001 NEPVSLQLT VVEGKNTPL SENLNFTLQN EHV⁹DAGAWRY QLIRKDGFR
 1051 LHN¹⁰VPKEQEL SDKLGAGET EAALTAQQA LAAKQQA¹¹EK NAQSLDALIA
 1101 AGRNATEKAE SVAEPARQAG GENAGIMQAE EEKKRVQADK D¹²TALAKQEA
 1151 ETRPATTAFF RARRARRDLP PQPQPQPQAP QRDLISRYAN SGLSEFSATL
 1201 NSVFAVQDEL DRVFAEDRRN AVWTS¹³GIRD¹⁴T KHYRSQDFRA YRQOTDLRQI
 1251 GMQKNLGSGR VGILF¹⁵SHNRT GNTFFDDGIGN SARLHAGAVF QYQIGRGFDI
 1301 GISAGAGFSS GSLSDGIRGK IRRRLVHYGI QARYRAGFGG FGIEPHIGAT
 1351 RYFVQKADYR YENVNIA¹⁶TPG LAENRYRAGI KADYSFKPQA HISITPYLSL
 1401 SYTDAASGVK RTRNV¹⁷AVLA QDFGKTRSAB WGVN¹⁸AEIKGF TL¹⁹SLHAAAAK
 1451 GPQLEAQHSA GIKLGYRW*

Underlined and double-underlined sequences represent the active site of a serine protease (trypsin family) and an ATP/GTP-binding site motif A (P-loop).

50 ORF1-1 and ORF1ng show 93.7% identity in 1471 aa overlap:

orf1-1.pep 10 20 30 40 50 60
 MKTTDKRTTETHRKA¹PKTGRIRFSPAYLAI²CLSPGILPQAWAGHTYFGIN³YQYRDPFEN
 orf1ng-1 MKTTDKRTTETHRKA¹PKTGRIRFSPAYLAI²CLSPGILPQARAGHTYFGIN³YQYRDPFEN
 55 10 20 30 40 50 60
 orf1-1.pep 70 80 90 100 110 120
 KKGFAVGAKDIEVYNKGGELVGGSMTKAPMIDFSVVSRRNGVAALVGGDQYIVSAHNGGYN
 60 70 80 90 100 110 120
 orf1ng-1 KKGFAVGAKDIEVYNKGGELVGGSMTKAPMIDFSVVSRRNGVAALVGGDQYIVSAHNGGYN

		130	140	150	160	170	180
	orf1-1.pep	NVDFGAEGRNPDQHRFTYKIVKRNNYKAGTKGHPYGGDYHMPRLHKFVTD AEPVENTSYM					
5	orf1ng-1	NVDFGAEGSNPDQHRFSYQIVKRNNYKAGTNGHPYGGDYHMPRLHKFVTD AEPVENTSYM					
		130	140	150	160	170	180
	orf1-1.pep	DGRKYIDQNNYPDRVRIGAGROYWRSDEDEPNNRESSYHIASAYSWLVGGNTFAQNGSGG					
10	orf1ng-1	DGWKYADLNKYPDRVRIGAGROYWRSDEDEPNNRESSYHIASAYSWLVGGNTFAQNGSGG					
		190	200	210	220	230	240
	orf1-1.pep	GTVNLGSEKIKHSPYGFLLPTGGSGFDGSGSPFMFIYDAQKQWLVLTGPNPYIGKSNGF					
15	orf1ng-1	GTVNLGSEKIKHSPYGFLLPTGGSGFDGSGSPFMFIYDAQKQWLVLTGPNPYIGKSNGF					
		250	260	270	280	290	300
	orf1-1.pep	QLVRKDWFYDEIFAGDTHSVFYEPHQNGKYSFNDNNNGTKINAKHEHNSLPNRLKTRTV					
20	orf1ng-1	QLVRKDWFYDEIFAGDTHSVFYEPHQNGKYFNDNNNGAGKIDAKHKHYSLPYRLKTRTV					
		310	320	330	340	350	360
	orf1-1.pep	QLFNVLSSETAREPVYHAAGGVNSYRPLNNGENISFIDKGGKELIITSINQAGAGGLYF					
25	orf1ng-1	QLFNVLSSETAREPVYHAAGGVNSYRPLNNGENISFIDKGGKELIITSINQAGAGGLYF					
		370	380	390	400	410	420
	orf1-1.pep	QGDFTVSPENNETWQAGAVHISEDSTVTWKVNGVANDRLSKIGKGTLLHVQAKGENQGSSIS					
30	orf1ng-1	EGNFTVSPKNNETWQAGAVHISEDSTVTWKVNGVANDRLSKIGKGTLLHVQAKGENQGSSIS					
		430	440	450	460	470	480
	orf1-1.pep	VGDGTVILDQADDQGGKQAFSEIGLVSGRGTVQLNADNQFNPDKLYFGFRGRLDLNGH					
35	orf1ng-1	VGDGKVIDQADDQGGKQAFSEIGLVSGRGTVQLNADNQFNPDKLYFGFRGRLDLNGH					
		490	500	510	520	530	540
	orf1-1.pep	SLSFHRIQNTDEGAMIVNHNQDKESTVTITGNKDIAITGNNNSLDSKKEIAYNGWFGKED					
40	orf1ng-1	SLSFHRIQNTDEGAMIVNHNQDKESTVTITGNKDITTTGNNNSLDSKKEIAYNGWFGKED					
		550	560	570	580	590	600
	orf1-1.pep	TTKTNGRLNLNVQPAEDRTLLSGGTNLNGNITQTNGKLFFSGRPTPHAYNHLNDHWSQ					
45	orf1ng-1	ATKTNGRLNLNVQPEEADRTLLSGGTNLNGNITQTNGKLFFSGRPTPHAYNHLGSGWSK					
		610	620	630	640	650	660
	orf1-1.pep	KEGIPRGEIWNDDWINRTFKAENFQIKGGQAVVSRNVAKVGDWHLNSHAQAVGVAPH					
50	orf1ng-1	MEGIPQGEIWNDDWIDRTFKAENFHIQGGQAVVSRNVAKVGDWHLNSHAQAVGVAPH					
		670	680	690	700	710	720
	orf1-1.pep	QSHTICTRSDWITGLTNCVEKTIITDDKVIASLTKTDISGNVPLADHAHNLTLGLATINGNL					
55	orf1ng-1	QSHTICTRSDWITGLTSCTEKTIITDDKVIASLSKTDIRGNVSLADHAHNLTLGLATINGNL					
		730	740	750	760	770	780
	orf1-1.pep	SANGDTRTYTVSHNATQNGNLSLVGNAQATFNQATLNGNTSASGNASFNLSHDHAVNGSGLT					
60	orf1ng-1	SAGGDTHYTVTRNATQNGNLSLVGNAQATFNQATLNGNTSASDNASFNLSNNAVNGSGLT					
		790	800	810	820	830	840
	orf1-1.pep	SANGDTRTYTVSHNATQNGNLSLVGNAQATFNQATLNGNTSASGNASFNLSHDHAVNGSGLT					
65	orf1ng-1	SAGGDTHYTVTRNATQNGNLSLVGNAQATFNQATLNGNTSASDNASFNLSNNAVNGSGLT					

		790	800	810	820	830	840
		850	860	870	880	890	900
5	orf1-1.pep	LSGNAKANVSHSALNGVNSLADKAVHFESSRFTGQISGGKDTALHLKDSSEWTLPSGT	EL				
	orf1ng-1	LSDNAKANVSHSALNGVNSLADKAVHFESSRFTGKISGGKDTALHLKDSSEWTLPSGT	EL				
		850	860	870	880	890	900
		910	920	930	940	950	960
10	orf1-1.pep	GNLNLDNATITLNSAYRHDAGAQTGSATDAPRRRRRRRRSLLSVPTTSVESRFNTLT					
	orf1ng-1	GNLNLDNATITLNSAYRHDAGAQTGSAADAPRRRRR---RSLLSVPTPTSAAESRFNTLT					
		910	920	930	940	950	960
15	orf1-1.pep	VNGKLNQGGTFRFMSELFGYRSDKILKLAESSEGYTTLAVNNTGNEPASEQLTVVEGKDN					
	orf1ng-1	VNGKLNQGGTFRFMSELFGYRSGKILKLAESSEGYTTLAVNNTGNEPVSLEQLTVVEGKDN					
		960	970	980	990	1000	1010
20							
	orf1-1.pep	KPLSENLNFTLQNEHVDAGAWRYQLIRKDGFERLHNPVKEQELSDKLGK-----					
	orf1ng-1	TPLSNENFTLQNEHVDAGAWRYQLIRKDGFERLHNPVKEQELSDKLGKAGETEALTA					
25		1020	1030	1040	1050	1060	1070
	orf1-1.pep	----EAKKQAEKDNQASLDALIAAGRDVAKTESVAEPARQAGGENGVIMQAESEKRVQ					
	orf1ng-1	QAQLAAKQAEKDNQASLDALIAAGRNATEKAEVSAEPARQAGGENAGIMQAESEKRVQ					
30		1080	1090	1100	1110	1120	1130
	orf1-1.pep	ADKDTALAKQREAEATRPATTAFPRARRARDLPQLQPQPQPORDLISRYANSGLSEFS					
	orf1ng-1	ADKDTALAKQREAEATRPATTAFPRARRARDLPQPQPQPQPQPORDLISRYANSGLSEFS					
35		1140	1150	1160	1170	1180	1190
	orf1-1.pep	ATLNSVFVAVQDELDRVFAEDRRNAVWTSIGIRDTHKYSQDFRAYRQQDTRLQIGMQKNLG					
	orf1ng-1	ATLNSVFVAVQDELDRVFAEDRRNAVWTSIGIRDTHKYSQDFRAYRQQDTRLQIGMQKNLG					
40		1200	1210	1220	1230	1240	1250
	orf1-1.pep	SGRVGILFSHNKTENTFFDDIGNSARLAHGAVFYQYQIGIDRPIYIGISAGAGFSSGSLSDGI					
	orf1ng-1	SGRVGILFSHNKTENTFFDDIGNSARLAHGAVFYQYQIGRFDIGISAGAGFSSGSLSDGI					
45		1260	1270	1280	1290	1300	
	orf1-1.pep	GGKIRRRVLHYGIQARYRAGGGFGIEPHIGATRYFVQKADYRYENVNVIATPGLAFNRYR					
	orf1ng-1	RGKIRRRVLHYGIQARYRAGGGFGIEPHIGATRYFVQKADYRYENVNVIATPGLAFNRYR					
50		1310	1320	1330	1340	1350	1360
	orf1-1.pep	AGTKADYSFKPAQHISITPYLSLSYTDAAAGKVRTRVNTAVLAQDFGKTRSAEWGVNAEI					
	orf1ng-1	AGTKADYSFKPAQHISITPYLSLSYTDAAAGKVRTRVNTAVLAQDFGKTRSAEWGVNAEI					
55		1380	1390	1400	1410	1420	
	orf1-1.pep	KGFTLSLHAAAAGFPQLEAQAHSAGIKLGYRWX					
	orf1ng-1	KGFTLSLHAAAAGFPQLEAQAHSAGIKLGYRWX					
60		1430	1440	1450			
	orf1-1.pep	KGFTLSLHAAAAGFPQLEAQAHSAGIKLGYRWX					
	orf1ng-1	KGFTLSLHAAAAGFPQLEAQAHSAGIKLGYRWX					
65		1440	1450	1460			

In addition, ORF1ng shows 55.7% identity with hap protein (P45387) over a 1455aa overlap:

SCORES Initl: 1104 Initn: 4632 Opt: 2680
 Smith-Waterman score: 5165; 55.7% identity in 1455 aa overlap

5	orf1ng-1.pep	MKTTDKRTTETHRKA	10	20	30	40	50	60
	p45387	KKFTVGRNLNLTACISL	10	20	30	40	50	60
10	orf1ng-1.pep	KGKFAVGA	70	80	90	100	110	120
	p45387	KGKFTVGAQNIK	70	80	90	100	110	120
15	orf1ng-1.pep	NVDFGAEGSNPDQHR	130	140	150	160	170	180
	p45387	DVDFGAEGNNPDQHR	130	140	150	160	170	180
20	orf1ng-1.pep	DGWKYADLNKYPDR	190	200	210	220	230	240
	p45387	NGSTYSDRTKYPER	190	200	210	220	230	240
25	orf1ng-1.pep	GTVNLGSEKIKHSP	250	260	270	280	290	300
	p45387	GYSYLGSDVRKAG	250	260	270	280	290	300
30	orf1ng-1.pep	QLVRKDWFYDEIF	310	320	330	340	350	360
	p45387	QLVRKSYF-DEIF	310	320	330	340	350	360
35	orf1ng-1.pep	QLFNVSLSETARE	370	380	390	400	410	419
	p45387	TLANMSLPLKEK	370	380	390	400	410	419
40	orf1ng-1.pep	FEQNFIVSFKNNE	420	430	440	450	460	470
	p45387	FEQNFIVSFNSN	420	430	440	450	460	470
45	orf1ng-1.pep	SVGDGKVI	480	490	500	510	520	530
	p45387	SVGDGKVI	480	490	500	510	520	530
50	orf1ng-1.pep	HSLSFHRIQNTDE	540	550	560	570	580	590
	p45387	HSLSFKRIQNTDE	540	550	560	570	580	590
55	orf1ng-1.pep	EKDATKTN	600	610	620	630	640	650
	p45387	ETDKKNHGR	600	610	620	630	640	650
60	orf1ng-1.pep	ETDKKNHGR	660	670	680	690	700	710
	p45387	ETDKKNHGR	660	670	680	690	700	710

		1380	1390	1400	1410	1420	1430
orfing-1.pep	AFNRYRAGIKADYSFKPAQHISITPYLSLSTDAAGKVRTRVNTAVLAQDFGKTRSAEW						
5	p45387	:: :: :: :: :: ::					
		1300	1310	1320	1330	1340	1350
	AFNRYNAGIRVDYTFPTDNISVLQYFFVNVYDQTTVNLTVLQQPFGRYWQKEV						
	orfing-1.pep	1440	1450	1460	1469		
10		GVNAEIKGFTLSLHAAAAGFPQLEAQSAGIKLGYRWX					
		:: :: :: :: :: :: ::					
	p45387	GLKAEILHFQISAFISKQSGSLGQKQNVGVKLYRW					
		1360	1370	1380	1390		

Based on this analysis, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

15 Example 78

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 65>:

	1	..AAGGTCGTGGC	AATTTCGTGGA	AGA.CCGCTG	CGTGGCGCTG	TGCCTGCCGA
	51	CAGTTTTTGAA	CCGACCGCGC	AAAAATTGAA	CGCTGTTTAG	GGGCGTGGG
20	101	CACCATTTT	GTTTTATGAA	GATCAAAATG	TCGTCAAAGG	TTTGCAGAG
	151	CAGTTCCTCTG	CTTATGCCGC	TAACTTCCCC	GTTTGGCGGg	ATCAGGCAA
	201	CGCGATGGTG	CAGTATGCCG	TTTGGACGAC	ACTTGCCGCG	GTCCGCGTAG
	251	GTGCAAACT	GCAACATTAC	AATCCTCTGC	CGATGCGCGC	GATGCGCAA
	301	CGGTGGAATA	TCCCCGAAAA	CTGGTTGTGT	CSCGCAAA	TGGTTATCGG
25	351	CGGTATTGAA	GGGCGGCGAG	GTGAAAAGAC	CTTTGAACCC	GTTCAGAAC
	401	GTTGAAAGT	GTCGGCGCA	TAA		

This corresponds to the amino acid sequence <SEQ ID 656; ORF6>:

	1	..KVWQFVEXPL	RAVVPADSFE	PTAQKLNLFK	AGAATILFYE	DQNVVKGLQE
	51	QFPAYAAANFP	VWADQANAMV	QYAVVMTLAA	VGVGANLQHY	NPLPDAAIAK
30	101	AWNIPENWLL	RAQMVIGGIE	GAAGEKTFEP	VAERLKVFGA	*

Further sequence analysis revealed a further partial DNA sequence <SEQ ID 657>:

	1	..CTGCGTGGCG	TCGTGGCTGC	CGACAGTTTT	GAACCGACCG	CGCAAAATT
	51	GAACCTGTTT	AAGGCGGGTG	CGGCAACCAT	TTTGTTTTAT	GAAGATCAAA
	101	ATGTCGTCAA	AGGTTTGACG	GAGCACTTCC	CTGCTATATG	CGCTAATCTC
	151	CCGCTTTGGG	CGGATCAGGC	AAACCGCATG	GTGCATATG	CGGTTGGAC
35	201	GACACTTGCC	CGCGTCGGCG	TAGTGCAGAA	CTCTGACAT	TACATCCCT
	251	TCGCCGATGC	GGCGATTGCC	AAACCGTGGA	ATATCCCCGA	AAACTGGTTC
	301	TTGCGGCGAC	AAATGGTTAT	CGCGGTATT	GAGGGGCGG	CGATGAA
	351	GACCTTTGAA	CCCGTGCAG	AACSTTTGAA	AGTGTTCGCG	CAGTAA

This corresponds to the amino acid sequence <SEQ ID 658; ORF6-1>:

	1	..LRAVVADSF	EPTAQKLNLF	KAGAATILFY	EDQNVVKGLQ	EQFPAYAAANF
	51	PVWADQANAM	VQYAVVMTLA	AVGVGANLQH	YNPLPDAAIA	KAWNIPENWL
40	101	LRAQMVIGGI	EGAAAGEKTFE	PVAERLKVFG	A*	

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

45 ORF6 shows 98.6% identity over a 140aa overlap with an ORF (ORF6a) from strain A of *N.meningitidis*:

			10	20	30
50	orf6.pep		KVWQFVEXPLRAVVADSF	EPTAQKLNLFK	
	orf6a	QIVEHAVLHTPSSENSQ	SARVVVLFGEHDKVQF	VEDALRAVVADSF	EPTAQKLNLFK

-396-

		40	50	60	70	80	90
5	orf6.pep	AGAATILFYEDQNVVKGLEQEFPAYAANFFVWADQANAMVQYAVWTTTLLAAVGVGANLQHY					
	orf6a	AGAATILFYEDQNVVKGLEQEFPAYAANFFVWADQANAMVQYAVWTTTLLAAVGVGANLQHY					
		100	110	120	130	140	150
10	orf6.pep	NPLPDAAIKAWNIPENWLLRAQMVI GGIEGAAGKTFEPVAERLKVFGAX					
	orf6a	NPLPDAAIKAWNIPENWLLRAQMVI GGIEGAAGKTFEPVAERLKVFGAX					
		160	170	180	190	200	

The complete length ORF6a nucleotide sequence <SEQ ID 659> is:

15	1	ATGACCCGTC	AATCTCTGCA	ACAGGCTGCC	GAAAGCGCCG	GTTCATTITA
	51	TTCGTTAAAT	AAAAATCTGC	CCGTCGGCAA	AGATGAARATC	GTCCAAATCG
	101	TCGAACACGC	CGTTTTCGAC	ACACCTCTCT	CGTTCAATTC	CCATCTGCC
	151	CGTGTGGTGC	TGCTGTTTGG	CGAAGAGCAT	GATAGGTGTG	GGCAATTTGT
20	201	CSAAGACGCG	CTGCGTCCG	TCGTGCGTGC	CGACAGTTT	GAACCGACCG
	251	CGCAAAAT	GAACTGTGTT	AAGCGGGTGC	CGCAACTACT	TTTGTTTTAT
	301	GAAGATCAAA	ATGTCGTCAA	AGGTTTGCG	GAGCAGTTCC	CTGCTATGCG
	351	CGCCAACTTT	CCGTTTGGG	CGGACCGAGC	GAACGGATG	GTGCAAGTATG
	401	CCGTTTGAC	GACACTTGCC	CGGTCGGCG	TAGGTGCAA	CCTGCAAACT
	451	TACAATCCCT	TGCCCGATGC	GGCGATTGCC	AAAGCGTGA	ATATCCCCGA
25	501	AACTGGTTG	TGCGCGCAC	AARTGGTTAT	CGGCGGTATT	GAAGGGGCGC
	551	CAGGTGAAA	GACCTTGAA	CCAGTTGCAG	AACGTTTGA	AGTGTTCGGC
	601	GCATAA				

This is predicted to encode a protein having amino acid sequence <SEQ ID 660>:

30	1	MTRQSLQQA	ESRRSIYSLN	KNLFVKGDEI	VQIVEHAVIL	TPSSFSNSQSA
	51	RVVVLFGEEH	DKVWQFVEDA	LRAVVPADSF	EPTAQKLNLF	KAGAATILFY
	101	EDQNVVKGLE	QEFPAYAANF	FWADQANAM	VQYAVWTTLA	AVGVGANLGH
	151	YNPLPDAIA	KAWNIPENWL	LRAQMVI	EGAAGEKTFE	PVAERLKVFG
	201	A*				

35 ORF6a and ORF6-1 show 100.0% identity in 131 aa overlap:

		50	60	70	80	90	100
	orf6a.pep	TPSSFSNSQSA	RVVVLFGEEHDKVWQFVEDALRAVVPADSF	EPTAQKLNLFKAGAATILEY			
	orf6-1			LRAVVPADSF	EPTAQKLNLFKAGAATILEY		
40					10	20	30
	orf6a.pep	EDQNVVKGLEQEFPAYAANFFVWADQANAMVQYAVWTTTLLAAVGVGANLQHYNPLPDAIA					
	orf6-1	EDQNVVKGLEQEFPAYAANFFVWADQANAMVQYAVWTTTLLAAVGVGANLQHYNPLPDAIA					
45		40	50	60	70	80	90
	orf6a.pep	KAWNIPENWLLRAQMVI GGIEGAAGKTFEPVAERLKVFGAX					
	orf6-1	KAWNIPENWLLRAQMVI GGIEGAAGKTFEPVAERLKVFGAX					
50		100	110	120	130		

Homology with a predicted ORF from *N.gonorrhoeae*

55 ORF6 shows 95.7% identity over a 140aa overlap with a predicted ORF (ORF6ng) from *N.gonorrhoeae*:

	orf6.pep	KVWQFVEXPLRAVVPADSF	EPTAQKLNLFK	30
	orf6ng	SNVSLDMSNPTVLRMLPLYIASLRRGAIYKWQFVEDALRAVVPADSF	EPTAQKLNLFK	64

```

orf6.pep    AGAATILFYEDQNVVKGLQEQFPAYAAFPVWADQANAMVQYAVVTTTAAVGVGANLQHY    90
             |||
orf6ng      AGRATILFYEDQNVVKGLQEQFPAYAAFPVWADQANAMVQYAVVTTTAAVGVGANLQHY    124
5  orf6.pep    NLPDAAIAKAWNIPENWLLRAQMVIIGGIEGAAGEKTFFPVAERLKVFGA    140
             |||
orf6ng      NLPDVAIAKAWNIPENWLLRAQMVIIGGIEGAAGEKVFEPVAERLKVFGA    174

```

The complete length ORF6ng nucleotide sequence <SEQ ID 661> was identified as:

```

10 1  ATGGCCGTTG  CGTCAATGT  CAGCTTGGAT  ATGTCCAATC  CTACGGTGTG
    51  ACGCATGGGA  TTACCCCTAT  ATATTGCGTC  CCTAAGAAGG  GCGCAATAT
   101  ATAAGGTGTG  GCAATTTGTC  GAAGACGCGC  TGCCTGCCGT  CGTGCTGCC
   151  GACAGTTTGT  AACCGACGCG  GCAAAAATTG  AAGCTGTTTA  AGGCGSGCGC
   201  GGCAACCAT  TTGTTTATG  AAGATCAAAA  TGTCGTCAAA  GTTTGCGAGG
   251  AGCAGTTCCC  TGCTTATGCC  GCCAATCTTC  CCGTTGSGGC  GSACAGGCG
   301  AACGCTATGG  TACAGTATGC  CGTCTGAGCG  ACATTPGCCG  CGGTCGGTG
   351  AGGTGCAAT  CTCGACCAT  ACACCCCTTC  GCCGATGTG  GCGATTGCTA
   401  AAGCGTGAA  TATTCGGAA  AACTGCGTGT  TGCGCGTGCA  AATGGTTATC
   451  GGTGGTATTG  AAGGCGGCG  agtgaaaaa  gtctttgaac  CCGTTCGgga
  501  acgtttgAAA  GTGTTGCGCG  CATAA

```

This encodes a protein having amino acid sequence <SEQ ID 662>:

```

1  MAVASNVSLD  MSNPTVLRMG  LPLYIASLRR  GAIYKVVQVF  EDALRAVVPA
51  DSFEPTAQKL  KLFKAGAATI  LFYEDQNVVK  GLQEQFPAYA  ANFPVWADQA
101  NAMVQYAVWT  TLAAVGAGAN  LQHYNPLPDV  AIAKAWNIPE  NWLLRAQMVI
151  GGIEGAAGEK  VFEPVAERLK  VFGA*

```

ORF6ng and ORF6-1 show 96.9% identity in 131 aa overlap:

```

30 orf6-1.pep    LRAVVPADSFEPTAQKLNLKFKAGATILFY
             10      20      30
orf6ng      PTVLRMGLPLYIASLRRGAIYKVVQVFEDALRAVVPADSFEPTAQKLKLFKAGATILFY
             20      30      40      50      60      70
35 orf6-1.pep    EDQNVVKGLQEQFPAYAAFPVWADQANAMVQYAVVTTTAAVGVGANLQHYNPLDVAIA
             40      50      60      70      80      90
orf6ng      EDQNVVKGLQEQFPAYAAFPVWADQANAMVQYAVVTTTAAVGVGANLQHYNPLDVAIA
             80      90      100     110     120     130
40 orf6-1.pep    KAWNIPENWLLRAQMVIIGGIEGAAGEKTFFPVAERLKVFGAX
             100     110     120     130
orf6ng      KAWNIPENWLLRAQMVIIGGIEGAAGEKVFEPVAERLKVFGAX
             140     150     160     170
45

```

It is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 79

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 663>

```

50 1  ..GGCTACAACT  ACCTGTTGCG  GCGCGGCAGC  CGCATCGCCA  ACTACCAAA
   51  CAACGGCRTC  CCGTTGCGG  ACGCGCTGGC  CGATAOGGGL  CAATGCCAAC
  101  ACCGCCGCT  ATGAGCGCGT  AGAAGTCGTG  CGCGCGGTGG  CGGGCTGTCT
  151  GGACGGCAGC  GCGGAGCCTT  CGGCCACGCT  CAATCTGGTG  CGCAACGCC
  201  TGACCCGCAA  GCCATTGTTT  GAGTCCGCG  CGCAGCGGgG  CRACCGCAA
  251  CATTTCCGGC  TGGACGCGGA  CGTATCGGGC  AGCCTGAACA  CCGAAG.crc
  301  rCTGCGGcGc  CGCCTGGTTT  CCAcCTTCGG  AOGCGCGCAC  TCGTGGCGGC

```

5
351 GGGGCGAAGC CAGCCGskAT GCCSAACTCT ACGGCATTTT GSAATACGAC
401 ATGCGCAGCG AAGCGCGCT CACGCGAAGC ATGGACTACC AGCAGCGGAA
451 AGAAGCGCGC GAGCGCGCGC TCAGCTACGC CGTGTACGAC AGCCAAAGTT
501 ATGCGCAGCG CTTCCGCGCGC AAGACCAACC CCGCCACAAA TTGGGGGAGAC
551 AGCCACACCC GTGCGCTCAA CTTGTTTCGC GGCATCGAAC ACOCGTTCAA
601 CCAAGACTGG AACTCAAAG CCGAATACGA CTAC..

This corresponds to the amino acid sequence <SEQ ID 664; ORF23>:

10
1 ..GYNLYFARGS RIANVOINGI PVADALADTG NANTAAYERV EVVRGVAGLL
51 DGTGEP SATV NLVRKRLTRK PLFEVRAEAG NRKHFLGLDAD VSGSLNTEXX
101 LRGRILVSTFG RGDSSRRRER SRXAELYGIL EYDIAPTQTV HAXMDYQQAN
151 ETADAPLSYA VYDSQGYATA FGPKNPATN WANSHHRLAN LFAGIEHRFN
201 QDWKLKAEYD Y..

Further work revealed the complete nucleotide sequence <SEQ ID 665>:

15
1 ATGACACGCT TCAAAATATTC CTGCTGTGTT GCCGCCCTGT TGCCCGTGTA
51 CGCGCAGGCC GATGTTTCTGT TTTACAGAGA CCCCAACCCG CAGGAAAGCG
101 CTGAATTGCC GACCATCACC GTTACCGCGC ACGCGACCCG GAGTTTCCAA
151 GAGCGCTACA CTGTTTCCGG CAGCGACACC CGCTGGGGC TGCCCATGAC
201 CTTGCGCGAA ATCCCGCAGA GCGTCAGCGT CATCATATCG CAACAAATGC
251 CGGACCAAAA CATCAAAAGC TCGACCGCGC CCTGTGTTGA GCGACCGGCG
20
301 ACCAGCGCGC AGATTATACGG CTCGACCGCG CGCGGTTACA ACTACTGTT
351 CGCGCGCGCG AGCCGCGATCG CCAACTACCA AATCAAGGCT ATCCCGCTTG
401 CCGACCGCTT GCGCGATACG GCGAATGCCA ACACCGCGCG CTATGAGCGC
451 CTAGAACTCG TCGCGCGCGT GCGGGGCGTG CTGACGCGCA CCGCGGAGCC
25
501 TTCCGCCACC GTCAATCTGG TGCGCAAAAC CTTGACCCCG AAGCAATTGT
551 TTGAAGTCCG CGCCGAAGCG GGCACCGCA AACATTTCGG GCTGGACCGG
601 GAGCTATCGG CGAGCGCTGAA CACCGAAGGC ACGCTGCGCG CGCGCTGTGT
651 TTCCACTTTC GGACGGCGGG ACTCGTGGCG GCGCGCGGAA CGCAGCCGCG
701 ATGCGGAACT CTACGGCATT TTGGAATACG ACATCGCACC GCAAAACCGC
30
751 GTCCAGCGAG GCATGGACTA CCAGCAGGCG AAGAAACCGC CGCAGCGCGC
801 GCTCAGCTAC CGCGTGTACG ACAGCGCAAG TTATGCGACC CCCTTGGCGC
901 CGAAAGACAA CCGCGCCACA AATTGGGCGA ACAGCGCCCA CCGTGCCTCT
951 AACTCTGTTC CGCGCATCGA ACACCGCTTC AACCAAGACT GGAAACTCAA
951 AGCCGAATAC GACTACACCC GCAGCGCGTT CGCGCAGGCC TAGGCGGTAG
35
1001 CAGGCGTGCT TTCCATCGAC CACAAACCCG CGCGCACGCA CTGATTTCCT
1051 GGTATTATGGC AGCGCGACCC GCGCACCCAC AGCGCCAGCG CTGATTGAT
1101 CGGCAAAATAC CGCCTCTTCG GCGCGACACA CGATTATATC GCGGATATCA
1151 ACGTTTACAA ATACGCCAGC AACAAATACG GCGGATATAC CATCATCCCC
40
1201 AAGCCATTTC CCAACGCCTA CGAATTTCCT CGCAGGGTGT CTAACCCGCA
1251 GCGTGCATCG TTTGCCAAA CCACTCCGCA ATACGCGACC AGGCGGCCAAA
1301 TCGCGCGGTA TCTCGCCACC CGTTTCCGCG CGCGCGACAA CTTTTCGCTG
1351 ATTTTGGGCG GAGGATACAC CGTTTACCGC ACGGGAGCT AGCAGCGCGC
1401 CACACAAGGC ATGACCTATG TGTCCGCAA CCGTTTACCC CCTTACACAG
1451 GCATCTGTGT GCACCTGACC GGCACCTGT CTCTTTACGG CTCGTACAGC
45
1501 AGCCTGTTCG TCCCGCAATC GCAAAAAGAC GAACACGGCA GCTACTGAA
1551 ACCCGTAACC GGCACAATC TGGAGCCGCG CATCAAAGCG GAATGGCTTG
1601 AAGGCCGCTC GAACGATCC GCGCGCGTGT ACCGCGCCCG TAAACAAAC
1651 CTCGCCACCG CAGCAGGACG GCACCGGAGC GGCACACCT ACTACGCGC
1701 CGCCAAACCA GCCAAAACCC ACGGCTGGGA AATCGAAGTC GCGGCCCGCA
50
1751 TCACGCCCGA ATGGCAGATA CAGGCAGGTT ACAGCCAAAG CAACACCCCG
1801 GACCAAGAGC GCAGCGCGCT GACCCCGCAC AGCTTACCCG AACCGAGCTT
1851 CAACCTCTTC ACTGCTTTCG AGCTTTCGCC CAGAGCCCGC AGCGCTTTCG
1901 CCAATCGCGC AGCGTTCGCG TCGCAGAGCG ARAACACAC CAGACCTGCC
1951 ACCTCTCGCA TCCCAACCCC CCGCGCCAAA GCGCGCGCG CGCACACAGC
55
2001 CCGCGCAAAA GCCTACGCGC TCGCGACAT CATGGCGGCT TACCGCTTCA
2051 ATCCGCGCGC GCAACTGTGC CTGAACGTGG ACAATCTGTT CACCAACAC
2101 TACCGCACCC AGCCCGACCG CCAAGCTAC GCGCGACTGC GGACAGTGAA
2151 CGCGCGCTTT ACCTATCGGT TTAAATAA

This corresponds to the amino acid sequence <SEQ ID 666; ORF23-1>:

60
1 MTRFKYLLF AALLEVYQA DVSVDPKP QESTLEPTIT VTADRTASSN
51 DGYTVSGTHT PLGLPMTLRE IPQSVSVITS QMRDNRNIT LDRALLQATG
101 TSRIQYSGDR AGYNLYFARG SRIANYQING IPVADALADT GNANTRAYER
151 VEVRGVAGL LDGTGEP SATV NLVRKRLTR KPLFEVRAEA GNRKHFLGLD
201 DVSGSLNTEG TLGRILVSTF GRGDSSRRER RSRDAELYGI LEYDIAPTQR

251 VHAGMDYQQA KETADAPLSY AVYDSQGYAT AFGPKDNPAT NWANSRHRAL
 301 NLFAGIEHRF NQDWKLKAEY DYTRSRFRQP YGVAVGLSID HNTAATDLIP
 351 GYWHADERTH SASVSLIGKY RLFGRHDLI AGINGYKYAS NKYGRSIIIP
 401 NAIPTNAYEFS RTGAYPQFAS FAQTIPQYGT RRQIGGYLAT RFRAADNLISL
 451 ILGGRYTRYR TGSYDSRTQG MTYVSANRFT PYTGIVFDLT GNLSLYGSYS
 501 SLFPVQSQKD EHGSLYLPVT GNNLEAGIKG EWLEGRNLAS AAVYRARNKN
 551 LATAAGRDPS GNTYYRAANQ AKTHGWELV GGRITPEWQI QAGYSQSKTR
 601 DQDGSRLNPD SVPERSEFKLF TAYHFAPEAP SGWTIGAGVR WQSETHTPDA
 651 TLRIPNPAK ARAADNSRQK AYAVADIMAR YRFNPRAELS LNVNLFNKH
 701 YRTQDRHSY GALRTVNAAF TYREK*

Computer analysis of this amino acid sequence gave the following results:

Homology with the ferric-pseudobactin receptor PupB of *Pseudomonas putida* (accession number P38047)

ORF23 and PupB protein show 32% aa identity in 205aa overlap:

Orf23 6 FARGSRINAYQINGIPVADALADTGNANTAAERYEVVRGVAGLLDGTGEP SATVNLVRK 65
 ++RG I NY+++G+P + L D + + A ++RVE+VRG GL+ G G PSAT+NL+RK
 PupB 215 WSRGFAIQYEVDPGVPTSTRL-DNYSQSMAMFDRVEIVRGATGLISGMGNPSATINLIRK 273
 Orf23 66 RLTRKPLFEVRAEAGNRKHFGLDADVSGSINTEXXLGRVLSTFXXXXXXXXXXXXAAE 125
 R T + + EAGN +G DVSG L +RGR V+ +
 PupB 274 RPTAEAAQSITGEAGNWDYRGTFDVSGLPTETGINRGFRVADYKTEKAWIDRYNQSSL 333
 Orf23 126 LYGILEYDIAPQTRVHAXMDYQQA KETADAPLSYAYVD--SQGYATAFGPKDNPATNWN 183
 +YGI E+D++ T + Y + D+PL + S G T N A +W+
 PupB 334 MYGITFDLSEDTLLTVGFSY--LRSDIDSPRLSGPLTRFSTGERTNLKRSINAAFDWSY 391
 Orf23 184 SHHRAALNLFAGIEHREFNQDWKLKAE 208
 + H + F IE + W K E
 PupB 392 NDHEQTSFPTSIEQQLGNWGSKIE 416

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF23 shows 95.7% identity over a 211aa overlap with an ORF (ORF23a) from strain A of *N.*

meningitidis:

35 orf23.pep 10 20 30
 GYNYLFARGSRINAYQINGIPVADALADTG
 orf23a QMRDNKIKALDRALLQATGTSRQIYSGSDRAGYNYLFARGSRINAYQINGIPVADALADTG
 90 100 110 120 130 140
 40 orf23.pep 40 50 60 70 80 90
 NANTAAYERVEVVRGVAGLLDGTGEP SATVNLVRKRLTRKPLFEVRAEAGNRKHFGLDAD
 orf23a NANTAAYERVEVVRGVAGLLDGTGEP SATVNLVRKRPTRKPLFEVRAEAGNRKHFGLGAD
 150 160 170 180 190 200
 45 orf23.pep 100 110 120 130 140 150
 VSGSINTEXXLGRVLSTFGRGDSWRRRRRSRXAEYLGILEYDIAPQTRVHAXMDYQQA
 orf23a VSGSINAEGLRGLVLSTFGRGDSWRQRRSRDLEYLGILEYDIAPQTRVHAXMDYQQA
 210 220 230 240 250 260
 50 orf23.pep 160 170 180 190 200 210
 ETADAPLSYAYVDSQGYATAFGPKDNPATNWNASHHRAALNLFAGIEHREFNQDWKLKAEYD
 orf23a ETADAPLSYAYVDSQGYATAFGPKDNPATNWNASHHRAALNLFAGIEHREFNQDWKLKAEYD
 270 280 290 300 310 320
 55 orf23.pep Y
 I
 60 orf23a YTRSRFRQPYGVAGVLSIDHNTAATDLIPGYWHADERTHSASVSLIGKYRFLGREHDLIA
 330 340 350 360 370 380

The complete length ORF23a nucleotide sequence <SEQ ID 667> is:

```

1   ATGACACGCT   TCAAATATTC   CCTGCTGTTT   GCGGCCCTGT   TGCCCGTGTGA
5   51   CGCGCAGGCC   GATGTTTCTG   TTTACAGACGA   CCCAAACCGT   CAGGAAGACA
101  101  CTGAATTGCC   GACCATCAAC   GTTACCGCGG   ACCGCAACCG   GAGTTCACAC
151  151  GACGGCTACA   CTGTTTCCGG   CACGCACACC   CGCGTCGGCG   TGCCCATGAC
201  201  CCTGGCGGAA   ATCCGCGAGA   GCGTCAGGGT   CATCACTATG   CAACAAATGC
251  251  GCGAGCAAAA   CNTCAAGCGG   CTCGACCGCG   CCTCTGTGCA   GCGGACCGCG
301  301  ACCGCGCGCC   AGATTACAGG   CTCGACCGCG   CTCGACCGCG   ACTACCTGTT
351  351  CGCGCGCGCT   AGCCGCATCG   CCAACTACCA   AATCAACGGC   ATCCCGTTTG
401  401  CGCGCGCGCT   GGCGGATACG   GGCATGCGCA   ACACCGCGCG   CTATGAGCGC
451  451  CTAGAATTCG   TCGCGCGCGT   GCGCGGGCTG   CTGGAGCGCA   CGGCGAGGCC
501  501  TTCCGCCACC   GTCAATCTGG   TGGCGAAACG   CCCGACCGCG   AAGCATTTGT
551  551  TTGAAGTCCG   CGCGGAAGCG   GGCACCGCCA   AACATTTCCG   GCTGGGCGCG
601  601  GACGCTATCG   GCAGCCTGAA   TGGCGAAGGC   ACGCTGCGCG   GCGCGCTGGT
15  651  TTCCACCTTC   GGACGCGCGG   ACTCGTGGCG   CGAGCGCGAA   CGCAGCGCGG
701  701  ATGCGGAACCT   CTACGGCATT   TTGGAATACG   ACATCGCACC   GCAAAACCGC
751  751  GTCCACGCG   GCATGGACCT   CACGACGGCG   AAGAAACCGC   CGCAGCGCGC
801  801  GCTCAGCTAC   GCGCTGTACG   ACAGCCTAAG   TTATGCCACC   GCCTTCGGCC
20  851  CGAAAGACAA   CCCCGCCACA   AATTGGGGCA   ACAGCGGCCA   CGGTGGCGTC
901  901  AACCTGTTTC   CGCGCATCGA   ACACGCGCTT   AACCAAGACT   GGAACCTCAA
951  951  AGCGGAATAC   GACTACACCC   GACGCGCGTT   CGCGACCGCC   TACGGCGTAG
1001 1001  CAGGCGTGCT   TTCCATCGAC   CACAACACCG   CCGCACCGCA   CTTGATTTCC
1051 1051  GGTATTGTCG   ACGCGGACCC   GCGGACCGCC   AGCGCGCGCG   TGCTCATTAAT
1101 1101  CGGCAAAATC   CGCGCTGTTC   GCGCGCAACG   CGATTTAATC   CGCGGATATCA
25  1151  ACGGTTACAA   ATACGCGCAG   AACAAATACG   CGCAGCGCAG   CATCATCTCC
1201 1201  AACGCGATTC   CCAAGCGCTA   CGAATTTTTC   CGCAGCGGTC   CCTACCGCGA
1251 1251  GCGTCATGCG   TTTGCCCAAA   CATCTCCGCA   ATACGCGCAA   ACGCGCGAAA
1301 1301  TCGGCGGCTA   TCTCGCACCC   CGTTTCCGCG   CGCGCGGCAA   CCTTTGCGTG
30  1351  ATACTCGGCG   GCAGATACAG   CGGTACCGCG   ACCGCGAGCT   ACGACGCGCG
1401 1401  CACACAAGCG   ATGACCTATG   TGTTCGCGAA   CCGTTTACCC   CCCTACACAG
1451 1451  GCATCTGTTT   GCACCTGACC   GGCACCTGTT   CGCTTTACGG   CTCGTACAGC
1501 1501  AGCCTGTTCG   TCCCGCAATC   GCAAAAAGAC   GAAACGCGCA   GCTACCTGAA
1551 1551  ACCGTAACCC   GGCACCAATC   TGGAGCGCGG   CATCAAAGGC   GAATGCGTTG
35  1601  AAGGCGGCTC   GAACGATCC   GCGCGCGGTG   ACCGCGCGCG   TAAACACAC
1651 1651  CTCGCCACCG   CAGCAGGACG   GCACCGGACG   GGCACAACCT   ACTACGCGCG
1701 1701  CGCCAACCAA   GCCAAAACCC   ACGGCTGGGA   AATCGAAGTC   GCGCGCGCGA
1751 1751  TCACGCCCGA   ATGGCAGATA   CAGGCGAGTT   ACAGCCAAAG   CAAAACCCGC
1801 1801  GACCAAGACG   GCAGCGCGCT   GAACCCCGAC   AGCGTACCGG   AACGCGAGCT
1851 1851  CAAACTCTTC   ACTGCGCTACC   ACTTTGCGCC   CGAAGCCGCC   AACGCGTGA
40  1901  CCAATCGCGC   AGCGGTGCGC   TGGCAGAGCG   AAACCCACAC   GACCCCTGCC
2001 2001  CGCGCAAAAC   GCGCGCAACC   CGCGCGCGCG   CGCAGCAACG
2051 2051  ATCCGCGCGC   CGAATCTGCG   CTGACCTGGG   ACAATCTGTT   CAACAAACAC
2101 2101  TACCGCACCC   AGCCGACCGC   CACGACGTAC   GCGGCACTGC   GGACAGTGAA
45  2151  CGCGCGGTTT   ACCTATCGGT   TTAATAAA

```

This encodes a protein having amino acid sequence <SEQ ID 668>:

```

1   MTRFKYSLLF   AALLPVYAQA   DVSVSDDPKP   QESTELPTIT   VTADRTASSN
5   51   DGYTVSGTHT   PLGLPMTLRE   IPQSVSVITS   QQMRDQNIKA   LDRALLQATG
101  101  TSRQIYGSDDR   AGYNYLFPARG   SRIANYPQIR   IPVADALADT   GNANTAAZER
151  151  VEVVRGVAGL   LDGTGEPSAT   VNLVRKRPTR   KPLEFVRREA   GNRKHGFLGA
201  201  DVSGLSLNAEG   TLRGLRVSTF   GRGDSWRQRE   RSRDAELYGI   LEYDIAPQTR
251  251  VHAGMDYQQA   KETADAPLSY   AVYDSQGYAT   AFGPKDNPAT   NWNASRRHAL
301  301  NLFAGIEHRFF   NQDWKLKAEY   DITRSFRFRQ   IGVAGVLSID   HNTAATDLIP
351  351  GYWHADPETH   SASVSLIGKY   RLFGREHDLI   AGINGYKYAS   NKYERSIIP
401  401  NAIPNAYEFS   RTGAVPOPAS   FQQTIPQYGT   RRGQIGYLAT   RFRADNLSL
451  451  ILGGRYSYVR   TGSYDSRRTQ   MITYYSANRET   PYTGIVFDLT   GNLISYVSLY
501  501  SLFVFPQSQKD   EHGYSYLPKPT   GNNLEAGLKG   ENLEGRLNAS   AAVYRARKNN
551  551  LATAAGRDPS   GNTYVRAAQQ   AKTHGWEIEV   GGRITPEWQI   QAGYSQSKTR
601  601  DDGDSRLNPD   SVFERSFKLF   TAYHFAPEAP   SGWTIGAGVR   WQSEHTIDFA
651  651  TLRIPNPAAK   ARAADNSRQK   AYAVADIMAR   YRFNPAEELS   LNVNLSNKH
701  701  YRTPQDRHSY   GALRTVNAAF   TYRFK*

```

ORF23a and ORF23-1 show 99.2% identity in 725 aa overlap:

```

                                10      20      30      40      50      60
orf23a.pep  MTRFKYSLLF AALLPVYAQA DVSVSDDPKP QESTELPTIT VTADRTASSN DGYTVSGTHT

```

	orf23-1	 MTRFKYSLFLFAALLPVYAQAADVSVSDDPQESTELPTTIVTADRTASSNDGYTVSGTHT 10 20 30 40 50 60
5	orf23a.pep	70 80 90 100 110 120 PLGLPMTLREIPQSVSVITSQQMRDQNIKALDRALLQATGTSRQYIGSDRAGYNLYFARG :
	orf23-1	70 80 90 100 110 120 PLGLPMTLREIPQSVSVITSQQMRDQNIKTLDRALLQATGTSRQYIGSDRAGYNLYFARG 70 80 90 100 110 120
10	orf23a.pep	130 140 150 160 170 180 SRIANYQINGIPVADALADTGNANTAAYERVEVVRGVAGLLDGTGEP SATVNLVRKRPT
15	orf23-1	130 140 150 160 170 180 SRIANYQINGIPVADALADTGNANTAAYERVEVVRGVAGLLDGTGEP SATVNLVRKRPT 130 140 150 160 170 180
	orf23a.pep	190 200 210 220 230 240 KPLFEVRAEAGNRKHFGLGADVSGSLNAGETLRGLVSTFGRGDSWRQRERSRDAELYGI
20	orf23-1	190 200 210 220 230 240 KPLFEVRAEAGNRKHFGLDADVSGSLNTEGTLRGLVSTFGRGDSWRRRERSRDAELYGI 190 200 210 220 230 240
	orf23a.pep	250 260 270 280 290 300 LEYDIAPOQTRVHAGMYQAKETADAPLSYAVYDSQGYATAFGPKDNPATNWNASRHRAL
25	orf23-1	250 260 270 280 290 300 LEYDIAPOQTRVHAGMYQAKETADAPLSYAVYDSQGYATAFGPKDNPATNWNASRHRAL 250 260 270 280 290 300
	orf23a.pep	310 320 330 340 350 360 NLFAGIEHRFNQDWKLEADYDTRSRFRQPYGVAGVLSIDHNTAATDLIPGYWHADPRTH
30	orf23-1	310 320 330 340 350 360 NLFAGIEHRFNQDWKLEADYDTRSRFRQPYGVAGVLSIDHNTAATDLIPGYWHADPRTH 310 320 330 340 350 360
	orf23a.pep	370 380 390 400 410 420 SASVSLIGKYRLFEGREHDLIAGINGKYASNKYGERSIIPNAIPNAYEFSTRTGAYPQPAS
35	orf23-1	370 380 390 400 410 420 SASVSLIGKYRLFEGREHDLIAGINGKYASNKYGERSIIPNAIPNAYEFSTRTGAYPQPAS 370 380 390 400 410 420
40	orf23a.pep	430 440 450 460 470 480 FAQTIPQYGTTRQIGGYLATRFRAADNLSLILGGYRTRYRTGSYDSRTQGMTYVSANRFT
45	orf23-1	430 440 450 460 470 480 FAQTIPQYGTTRQIGGYLATRFRAADNLSLILGGYRTRYRTGSYDSRTQGMTYVSANRFT 430 440 450 460 470 480
	orf23a.pep	490 500 510 520 530 540 PYTGIVFDLTGNLSLYGSYSLEVPQSQKDEHGSYLKPVTTGNLLEAGIKGEWLEGRINAS
50	orf23-1	490 500 510 520 530 540 PYTGIVFDLTGNLSLYGSYSLEVPQSQKDEHGSYLKPVTTGNLLEAGIKGEWLEGRINAS 490 500 510 520 530 540
	orf23a.pep	550 560 570 580 590 600 AAVYRARKNNLATAAGRDPSGNTYYRAANQAKTHGWEIEVGGRIITPEWQIQAGYSQSKTR
55	orf23-1	550 560 570 580 590 600 AAVYRARKNNLATAAGRDPSGNTYYRAANQAKTHGWEIEVGGRIITPEWQIQAGYSQSKTR 550 560 570 580 590 600
	orf23a.pep	610 620 630 640 650 660 DQDGSRLNDPSVPERSFKLFTAYHFAFEAPSGWTIGAGVRWQSETHDTPATLRIPNFAAK
60	orf23-1	610 620 630 640 650 660 DQDGSRLNDPSVPERSFKLFTAYHFAFEAPSGWTIGAGVRWQSETHDTPATLRIPNFAAK 610 620 630 640 650 660
65	orf23a.pep	670 680 690 700 710 720 ARAANSRQKAYAVADIMARYRFPRAELSLNVNLFNKHYRTQPDHRSYGALRTVNAAF
	orf23-1	670 680 690 700 710 720 ARAANSRQKAYAVADIMARYRFPRAELSLNVNLFNKHYRTQPDHRSYGALRTVNAAF 670 680 690 700 710 720

orf23a.pep TYRFFKX
 |||||
 orf23-1 TYRFFKX

Homology with a predicted ORF from *N.gonorrhoeae*

ORF23 shows 93.4% identity over a 211aa overlap with a predicted ORF (ORF23.ng) from *N. gonorrhoeae*:

orf23.pep	GYNVLFARGSRIANYQINGIPVADALADTGNANTAAYERVEVVRGVAGLLD	51
orf23ng	SAVDACRIPGYNVLFARGSRIANYQINGIPVADALADTGNANTAAYERVEVVRGVAGLPD	60
orf23.pep	GTGEPSTVNLVVRKRLTRKPLFEVRAEAGNRKHFGLDADVSGSLNTEXXLRGLRVSTFGR	111
orf23ng	GTGEPSTVNLVVRKHPTRKPLFEVRAEAGNRKHFGLGADVSGSLNABGTLRGLRVSTFGR	120
orf23.pep	GDSWRRERSRKAELYGILEYDIAPQTRVHAXMDYQQAETADAPLSYAVYDSQGYATAF	171
orf23ng	GDSWRQLERSDAELYGILEYDIAPQTRVHAGMDYQQAETADAPLSYAVYDSQGYATAF	180
orf23.pep	GPKNDPATNWNASHHRLNLFAGIEHRFNQDWKLKAEYDY	211
orf23ng	GPKNDPATNWSNRNALNLFAGIEHRFNQDWKLKAEYDYTRSRFRQPYGVAGVLSIDHS	240

The ORF23ng nucleotide sequence <SEQ ID 669> is predicted to encode a protein comprising amino acid sequence <SEQ ID 670>:

1	SAVDACRIPG	YNYLFARGSR	IANYQINGIP	VADALADTGN	ANTAAYERVE
5	VVRGVAGLPD	GTGEPSTVNL	VVRKHPTRKPL	FEVRAEAGN	RKHFGSLGADV
10	SGSLNABGTL	RGLRVSTFGR	GDSWRQLERS	RDABLYGILE	YDIAPQTRVH
15	AGMDYQQAKE	TADAPLSYAV	YDSQGYATAF	GPKNDPATNW	SNRNRNALNL
20	FAGIEHRFNQ	DWKLKAEYDY	TRSRFRQPYG	VAGVLSIDHS	TAATULIPGY
25	WHADPRTHSA	SMSLTGKYRL	FGREHLLIAG	INGKYASNNK	YGERSLIPMA
30	IPNAYEFSRT	GAYPQSSSEA	QTIPQDTRR	QIGGYLATRF	RAADNLPLL
35	GGYRSRYRAG	SYNSRTQGMT	VYSANRFTFY	TGIVFOLDTN	LSLYGYSSSL
40	FVPQLQKDEH	GSYLKPYTCN	NLEADIKGEW	LEGLRLNASAA	VYRARKNNILA
45	TAAGRDSQSN	TYYRAANQAK	THGWEIEVGG	RTPEWQIQQA	GYSQSKPRDQ
50	DGSRINPDSV	PERSFKLFYA	YHLAPEAPSG	RTIGAGVRRQ	GETHTDPAAL
55	RIPNPAAKAR	AVANSRQKAY	AVADIMARYR	FNPRTELSLN	VDNLFNKHYR
60	TQPDHRYSYA	LRTVNAAFTY	REFK*		

Further work revealed the complete nucleotide sequence <SEQ ID 671>:

1	ATGACACGCT	TCAAATACTC	CCTGCTTTTT	GCCGCGCTCG	TACCCGTGTA
5	CGCGCAGGCC	GATGTTTCTG	TTTCAGACGA	CCCCAACC	CAGGAAGACA
10	CCGAATTGCC	GACCATCAC	GTTACCGCCG	ACCGCACCGC	GAGTTCACAC
15	GACGGCTACA	CGGTTTCCGG	CACGCACACC	CGGTTGGGCG	TGCCCATGAC
20	CCTGCGCGAA	ATCCCGCAGA	GCGTCAGCGT	CATCACATCG	CAACAATGCG
25	GCAGCCAAAA	CATCAAAACG	CTCGACCCGG	CCCTGTTCGA	GGCGACCCGG
30	ACGACGCCGC	AGATTATCGG	CTCCGACCGC	CGCGCTACCA	ACTACCTGTT
35	CCGCGCGCGC	ACCGCATCGC	CCACATACCA	AAATCACAGC	ATCCCGCTGG
40	CGCGCGCGCT	GGCGCATACG	GGCAATGCCA	ACACCGCGCC	CTATGACGCG
45	CTAGAAGTCG	TGCGCGGCGT	GGCGGGGCTG	CGCGAGGCCA	CGGCGGAGCC
50	TTCTGCCACC	GTCATCTCGG	TACGCAACAA	CCCCACCGCC	AAGCCATTGT
55	TTGAAGTCCG	CGCGGAAGCC	GGCAACCGCA	AACATTTCGG	GCTGGGCGGG
60	GACGTATCGG	CGAGCCTGAA	CGCGGAAGCC	ACGCTGCGCG	CGCGCCTGCT
65	TTCCACCTTC	GGACGCGCGG	ACTCGTGCGG	CGAGCTCGAA	CGCAGCCGGG
70	ATGCCGAACT	CTACGGCATTT	TTGGAATACG	ACTCGCACCC	GCAACCCCGC
75	GTCACGCGAG	CGATGGACTA	CCAGCAGGCG	AAAGAACCGC	CAGACCGCGC
80	GCTCAGCTAC	CGCGTGTACG	ACAGCCAAAG	TTATGCCACC	CGCTTCGGCC
85	CAAAAGACAA	CCCCGCCACA	AATTGGTCGA	ACAGCCGCAC	CCGTGCGCTC
90	AACCTGTGTC	CGGGCATAGA	ACACCGCTTC	AACCAAGACT	GGAAACTCAA
95	AGCCGAATAC	GACTACACCC	GTAGCCGCTT	CCGCGAGCCC	TACGGTGTGG
100	CAGCGGTACT	TTCCATCGAC	CACAGCAGTG	CCGCGACCGA	CTGATTCCCG

5
10
15
20

```

1051 GGTATTATGGC ACGGcgatccc GCGCACCAC AGCGCCAGCA TGTCATTGAC
1101 CGCGAANTAC CgcctGTTGCG GCGCGGAGCA CGATTTAATC GCGGGTATCA
1151 ACGGCTACAA ATACGCCAGC AACAATAACG GCGAAGCAGC CATCATCTCC
1201 AACGCCATTCC CCAACGCCCTA CGAATTTTCC CGCAGCGGCG CCTATCCGCA
1251 GCCATCATCG TTTGCCCAA CCATCCCGCA ATACGACACC AGGCGGCAAA
1301 TCGGCGGCTA TCTCGCCACC CGTTTCCGCG CGCGCGACAA CCTTTCGTGT
1351 ATACTCTGGCG GCAGATACAG CGCTACC GCAGGCAGCT ACACACAGCG
1401 CACACAAGGC ATGACCTATG TGTCTGCCAA CGTTTACCC CCTACACAG
1451 GCATCGTGT CGATCTGACC GGCAACCTGT CGCTTTACGG CTGCTACAGC
1501 AGCCTGTTTC TCCCGCAATT GCAAAAAGAC GAACACGGCA GCTACCTGAA
1551 ACCCGTAACC GGCAACAMTC TGGGAAGCCGA CATCAAAGGC GAATGGCTTG
1601 AAGGGCGTCT GAACGCATCC GCGCCGTGT ACCGCGCCGG TAAAAACAA
1651 CTGCGCCACG CAGCAGGACC OGACAGAGC GGCAACACCT ACTATCGCGC
1701 CGCGAACCAA GCCAAAGACC ACGGCTGGGA AATCGAAGTC GCGCGCCGCA
1751 TCACGCGCGA ATGGCAGATA CAGGCAGSCT ACAGCCAAAG CAACACCCGC
1801 GACCAAGAGC GCAGCGCGCT GAACCCCGAC AGGCTACCGC AACCGACTTT
1851 CAAATCTTTC ACGCGCTACC ACTTAGCCCG CGAAGCCCGC AGCGCGCGGA
1901 CCATcggTGC GGGTGTGCGC GCGCGCGGCG AACCACACAC GCACCCAGCC
1951 GCGCTCGCA TCCCAACGCC CGCGCGCAAA GCCCGCGCGC TCGCAACAG
2001 CGCCAGAA GCGCTACGCC TCGCGACATC AACCACACAC GCACCGTTCA
2051 ATCCGCGCAC CGAATCTGTG CTGAACCTGG ACAACCTGTT CAACAAACAG
2101 TACCGCACCC AGCCCGACCG CCACAGCTAC GCGCGACTGC GGACAGTGAA
2151 CGCGCGCTTT ACCTATCGGT TTAATAAA

```

This corresponds to the amino acid sequence <SEQ ID 672; ORF23ng-1>:

25
30
35
40

```

1 MTRFKYSLLF AALLPVYAQA DVSVDSPK PQESTELPIT VTADRTASSN
51 DGYTVSGTHT PFGLPMTLRE IPQSVSVITS QMRDQNIKT LDRALLQATG
101 TSRIQYIGSDR AGYNLFFARG SRIANYQING IPVADALADT GNANTAAYER
151 VEVVRGVAGL PDGTGEP SAT VNLRKHPT RKLFEVRAEA GNRKHPLGGA
201 DVSGSLNAEG TLRGLRVSTF GRGDSWRGLE RSRDAELYGI LEYDIAPQTR
251 VHAGMDYQQA KETADAPLSY AVYDSQGYAT AFGPKDNPAT NWSNRRL
301 NLFAGIEHRF NQDWKLKAEY DYTRSRFRQP YGVAGVLSID HSTAATDLPI
351 GYWHADPRTH SASMSLTGKY RLFGREHDLI AGINGKYKAS NKYGRSLIP
401 NALPNAYEFS RTGAYPQSS FAQTIPQVDT ERQIGYGLAT RFAADNLSL
451 ILGGRYSRYR AGSYNSRFG HGYVSNRFT PFTGIVFDLT GNLISYGSYS
501 SLFVQLQKO EHSYVILKPT GNNLEADIKT EWLGRINAS AAIVYRARN
551 LATAAGRDQS GNTYBRAANQ AKTHGWIEIV GGRITPEWQI QAQYSQSKPR
601 QDQGSRLNPD SVPEFSFKLF TAYHLAEP AGSRTIGAGVR RQGETHTDPA
651 ALRIPNPAK ARAVANSRQK AYAVADIMR YRFPNPELS LNVNLFNKH
701 YRTQPDHSY GALRTVNAAF TYRFK*

```

40 ORF23ng-1 and ORF23-1 show 95.9% identity in 725 aa overlap:

45
50
55
60
65

```

10 20 30 40 50 60
orf23-1.pep MTRFKYSLLFAALLPVYAQADVSVDDPKPQESTELPITVTADRTASSNDGYTVSGTHT
orf23ng-1 MTRFKYSLLFAALLPVYAQADVSVDDPKPQESTELPITVTADRTASSNDGYTVSGTHT
10 20 30 40 50 60
70 80 90 100 110 120
orf23-1.pep PLGLPMTLREIPQSVSVITSQMRDQNIKTLDRLALQATGTSRIQYIGSDRAGYNLFFARG
orf23ng-1 PFGLPMTLREIPQSVSVITSQMRDQNIKTLDRLALQATGTSRIQYIGSDRAGYNLFFARG
70 80 90 100 110 120
130 140 150 160 170 180
orf23-1.pep SRIANYQINGIPVADALADTGNANTAAYERVEVVRGVAGLLDGTGEP SATVNLRKHPT
orf23ng-1 SRIANYQINGIPVADALADTGNANTAAYERVEVVRGVAGLPDGTGEP SATVNLRKHPT
130 140 150 160 170 180
190 200 210 220 230 240
orf23-1.pep KPLFEVRAEAGNRKHFGLDADVSGSLNAGETLRGLRVSTFGRGDSWRRRSRDAELYGI
orf23ng-1 KPLFEVRAEAGNRKHFGLDADVSGSLNAGETLRGLRVSTFGRGDSWRQLSRDAELYGI
190 200 210 220 230 240
250 260 270 280 290 300
orf23-1.pep LEYDIAPQTRVHAGMDYQQA KETADAPLSYAVYDSQGYATAFGPKDNPATNWSNRRL

```

	orf23ng-1	 LEYDIAFQTRVHAGMDYQQAETADAPLSYAVDYSQGVATAFGPKDNPNATWSNSRNRAL 250 260 270 280 290 300
5	orf23-1.pep	310 320 330 340 350 360 NLFAGIEHRFNQDWKLKAEYDYTRSRFRQPYGVAGVLSIDHNTAATDLIPGYWHADPRTH orf23ng-1 NLFAGIEHRFNQDWKLKAEYDYTRSRFRQPYGVAGVLSIDHNTAATDLIPGYWHADPRTH 310 320 330 340 350 360
10	orf23-1.pep	370 380 390 400 410 420 SASVSLIGKYRLFGREHDLIAGINGYKYASNKYGERSTIIPNAIPNAYEFSRTGAYPQPAS orf23ng-1 SASMSLTGKYRLFGREHDLIAGINGYKYASNKYGERSTIIPNAIPNAYEFSRTGAYPQPAS 370 380 390 400 410 420
15	orf23-1.pep	430 440 450 460 470 480 FAQTIPQYTRRQIGGYLATRFRAADNLSLLGGRYTRYRTGYSYDSRTQGMITYVSANRRT orf23ng-1 FAQTIPQYTRRQIGGYLATRFRAADNLSLLGGRYTRYRTGYSYDSRTQGMITYVSANRRT 430 440 450 460 470 480
20	orf23-1.pep	490 500 510 520 530 540 PYTGIVFDLTGNLSLYGSYSSLFVPQSKQDEHGSYLLKPVGTNNLEAGIKGEWLEGRINAS orf23ng-1 PYTGIVFDLTGNLSLYGSYSSLFVPQSKQDEHGSYLLKPVGTNNLEADIKGEWLEGRINAS 490 500 510 520 530 540
25	orf23-1.pep	550 560 570 580 590 600 AAVYRARKNNLATAAGRDPSGNTYYRAANQAKTHGWEIEVGGRIPTPEWQIQAGYSQSKTR orf23ng-1 AAVYRARKNNLATAAGRDQSGNTYYRAANQAKTHGWEIEVGGRIPTPEWQIQAGYSQSKPR 550 560 570 580 590 600
30	orf23-1.pep	610 620 630 640 650 660 DQDGSRLNPDSVPERSFKLFTAYHFAPEAPSGWTIGAGVRWQSEHTDPATLRIPNPAK orf23ng-1 DQDGSRLNPDSVPERSFKLFTAYHFAPEAPSGRTIGAGVRWQSEHTDPATLRIPNPAK 610 620 630 640 650 660
35	orf23-1.pep	670 680 690 700 710 720 ARAADNSRKAYAVADIMARYRFPNRAELSLNVNLFNKHYRTQPDHRSYGALRTVNAAF orf23ng-1 ARAVANSRKAYAVADIMARYRFPNRAELSLNVNLFNKHYRTQPDHRSYGALRTVNAAF 670 680 690 700 710 720
40	orf23-1.pep	TYRFFKK orf23ng-1 TYRFFKK

In addition, ORF23ng-1 shows significant homology with an OMP from *E.coli*:

55	sp P16869 FHUE_ECOLI Outer-MEMBRANE RECEPTOR FOR Fe(III)-COPROGEN, Fe(III)-FERRIOXAMINE B AND Fe(III)-RHODOTRULIC ACID PRECURSOR >gi 1651542 gnl PID d1015403 (D90745) Outer membrane protein FhuE precursor [Escherichia coli]
60	>gi 1651545 gnl PID d1015405 (D90746) Outer membrane protein FhuE precursor [Escherichia coli] >gi 1787344 (AE000210) outer-membrane receptor for Fe(III)-coprogen, Fe(III)-ferrioxamine B and Fe(III)-rhodotruclic acid precursor [Escherichia coli] Length = 729 Score = 332 bits (843), Expect = 3e-90 Identities = 228/717 (31%), Positives = 350/717 (48%), Gaps = 60/717 (8%)
65	Query: 38 TITVTPADRTASSN--DGYTYSGTHTFPGFLMTREIPQSVSVITSSQQRMDNIKTLDRAL 95 T+V TA+ +Y+V+T +MT R+IPQSV+++ Q+M DQ ++TL+ + Sbjct: 43 TVIVEGSATAPDGDENDYSVTSTSGATKMQMTRDIPQSVITVSQQRMEDQQLQTLGEVM 102
	Query: 96 LQATGTSRQIYGS DRAGYNLYLFARGSRANIYQINGIP-----VADALADTGNANTAA 147 G S+ SDR A Y ++RG +I NY ++GIP + DAL+D A Sbjct: 103 ENTLLGSKSQADSRLALY---YSRGFQIDNVMVDCIPYFESRWNLGDALSDM-----AL 154

Query: 148 YERVEVVRGAGLPGDTGCEPSATVNLVRKHPTKPLF-EVRAEAGNRKHFGLGADVSSGL 206
 +ERVEVVRG GL GTC PSA +N+VRKH T + +V AE G+ AD+ L
 Sbjct: 155 FERVEVVRGATGLMTGTGNPSAANMVRKHATSREFFKGDVSAEYGSWNKERYVADLQSLP 214

Query: 207 NAEGTLRGLVSTFGRGDSWRQLRSRDAELYGILEYDIAPQTRVHAGMDVQQAKETADA 266
 +G +R R+V + DSW S GI++ D+ T + AG +YQ+ +
 Sbjct: 215 TEDGKIRAIRVGGYQNNDSWLDRYNSEKTFSSGIVDADLGLTTLASAGVEQRIQDVSPT 274

Query: 267 PLSYAVYDSQGYATAFGPKDNPATNWSNRRNALNLFAGIEHRFPQDWKLAQYDITRSR 326
 +++ G + ++ ++ A +W+ + +F ++ +F W+ ++
 Sbjct: 275 WGGLPRWNTDGSNSYDRARSTAPDWAYNDKEINKVFMTLKQQFADTQWATLNATHSEVE 334

Query: 327 F--RQPYGVAGVLSIDHSTAA--TDLIPGY-----WHADPRTHTSA-SMSLTGKYRLFG 374
 F + Y A V D ++ PG+ W++ R A + + G Y LFG
 Sbjct: 335 FDSKMYVDAYVKNADGMLVGPYSNYGPGFDYVGGTGWNSGKRRVDALDLFADGSEYELFG 394

Query: 375 REHDLIAGINGKYKASNKYGER--SIIPNAIFNAYEFSRTGAYPQPSPFAQTIQYDTRR 432
 R+H+L+ G Y +N+Y +I P+ I + Y F+ G +PQ Q++ Q DT
 Sbjct: 395 RQHNLMTFG--GSYSQNNRYFSSWANIFFDEIGSFYNFN--GNFPQTWSPQSLAQDDTH 451

Query: 433 QIGGYLATFRFADNLILGGRYSRYRAGSYNSRTQGMTY-VSANRFTPYTGIVFDXXX 491
 Y ATR AD L LILG RY+ +R + +TY + N TPY G+VFD
 Sbjct: 452 MKSLYATRVTLADPLHLILGARYTNWRVDT-----LTYSMKHNHTPYAGLVFDIND 504

Query: 492 XXXXXXXXXXFXVFPQLQKDEHGSYLKPVNTGNLEADIKGEWLEGRNLNAAVYRARKNNL 551
 F PQ +D G YL P+TGN E +K +W+ RL + A++R ++N+
 Sbjct: 505 NWSTYASYSIFQPQNDRDSSGKYLAPITGNLYELGLKSDWMNRLTLTIAIFRIEQDNV 564

Query: 552 ATAAGR---DQSGNTYYRAANQAKTHGWEIEVGGRTPEWQIQAGYSQSKPRDQDGSRLN 608
 A + G +G T Y+A + + G E E+ G IT WQ+ G ++ D +G+ +N
 Sbjct: 565 AQSTGTPIPGSNGETAYKAVDGTVSKGVEFELNGAITDNQLTFGATRYAEDNMEGNAVN 624

Query: 609 PDSVPERSFKFLTAYHLAFAEAPSGRTIGAGVVRQGETHTDPAALRIPNPAKARAVANSR 668
 P ++P + K+FY L P F T+G GY Q +T D P RA
 Sbjct: 625 P-NLERTTYKMFYSYRL-PVME-LTVGGGVNWQNRVYTDIV-----TPYGTFFRA-----E 672

Query: 669 QKAYAVADIMARYFRNPTLESLNVNDLNFKNHYRTQPRH-SYGALRTVNAFTYRF 724
 Q +YA+ D+ RY+ L NV+NLF+K Y T + YG R + TY+F
 Sbjct: 673 QGSYALVDLTRYQVTKNLSLQGNVNNLFDKTYDITNVEGSIVYGTPRNFSITGYQF 729

Based on this analysis, it was predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF23-1 (77.5kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 15A shows the results of affinity purification of the His-fusion protein, and Figure 15B shows the results of expression of the GST-fusion in *E.coli*. Purified His-fusion protein was used to immunize mice, whose sera were used for Western blot (Figure 15C) and for ELISA (positive result). These experiments confirm that ORF23-1 is a surface-exposed protein, and that it is a useful immunogen.

Example 80

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 673>:

```

1   ATGCGCACGG CAGTGGTTTT GCTGTGTATC ATGCCGATGG CGGCTTCGTC
5   GGCAATGATG CCGGAAATGG TGTGCGCGGG CGTGTGCGCG GGAACGGCAA
101 TCATATCCAA GCCGACCGAA CAAACGGCGG TCATGGCTTC GAGTTTGTCC

```

151 AGCGTCAGcA CGCCTGCTTC GCGGcGgCa ATCATACCTT CGTCTTCGGA
 201 AACGGGATA AACCGCCAC TCAACCCCG GACCGCGCTG GAACGCATCA
 251 TCGCGCTTTT TTTCACGGCA TCGTTCAGCA ATGCCAAAGC TCGTGTGTGTG
 301 CCGTGCCTAC CGCAGACGCT CAAGCCCATT TnTTCAGAA TGCGTGCCAC
 351 TnASTCGCCG ACGGGG..

This corresponds to the amino acid sequence <SEQ ID 674; ORF24>:

1 MRTAVVLLLI MPMAASSAMM PEMVCAGVSP GTAILSKPTE QTAVMASSLS
 51 SVSTPASAAA IIPSSSETGI NAPLKPPTAL EAIMPPFFTA SFSNAKAADV
 101 PCVPQTLKPI XSRMRATXSP TG..

Further work revealed the complete nucleotide sequence <SEQ ID 675>:

1 ATGCGCACGG CAGTGGTTTT GCTGTTGATC ATGCCGATGG CGGCTTCGTC
 51 GGCAATGATG CCGGAAATGG TGTGCGGGGG CGTGTGCGCG GGAACGGCAA
 101 TCATATCCAA CGCAGCCGAA CAAACGGCGG TCGTGGCTTC GAGTTTGTCC
 151 AGCGTCAGCA CGCCTGCTTC GCGGGCGGCA ATCATACCTT CGTCTTCGGA
 201 AACGGGATA AACCGCCAC TCAACCCCG GACCGCGCTG GAAGCCATCA
 251 TCGCGCTTTT TTTCACGGCA TCGTTCAGCA ATGCCAAAGC TGCTGTGTGTG
 301 CCGTGCCTAC CGCAGACGCT CAAGCCCATT TCTTCAAGAA TGCGTGCCAC
 351 TGAGTCGCGC ACGGCGGGGG TCGGCGCCAG CGACAGTGTG AGAATACCAA
 401 ACGGGATATT CAGCATTTT GAGCGTTTCC GCGCGATGAG TTGCGCGACG
 451 CGGATATTG TCAAGCAGT TTCTTCACCT ACTTCGCGAA TCTCGCTCAA
 501 TGTCTTGTGA TCTGAATTTT CCAACGCGGC TTTTACGACA CTTGGGCGGG
 551 ATACGCGCAG ATTGATAACG GCATCCGCTT CGCCCGAAC ATGAACGCGG
 601 CCGCGCCATAA ACGGGTTGTC TTCCACCGCG TTGCAGAAC CGACAATTTT
 651 AGCGCAGCCG AACCTTTCGG GCGTGTATTC CGCCGTGCGT TTGACGGTTT
 701 CGCCGCGCAG CTTGACCCGA TCCATATTGA TACCGGCACG CGTACTGCGC
 751 ATATTGTATG AGCTGCACAC AATATCGGTA GTCTTATCAT CTTCCGGAAT
 801 GGAGCGGATT AACACCTCAT CCGAAGCGCA CATCCCTTTT TGCACCAAGC
 851 CGGAAAAACC GCGGATAAAA GACACACCGA TGCGTTTGGC AGCTTTATCC
 901 AAAGTTTGCG CCACGCTGAC GTAA

This corresponds to the amino acid sequence <SEQ ID 676; ORF24-1>:

1 MRTAVVLLLI MPMAASSAMM PEMVCAGVSP GTAILSKPTE QTAVMASSLS
 51 SVSTPASAAA IIPSSSETGI NAPLKPPTAL EAIMPPFFTA SFSNAKAADV
 101 PCVPQTLKPI SSRMRATESP TAGVGASDKS RIPNGIFSI EASRPMSSST
 151 RVILKAVFFT TSATSVNVVA SEFSNAFFT PGPTPTLIT ASASPEP*NA
 201 FAINGLSSTA LQNTTILAQP KESGVISAVR LEVSPASLTA SILIPARVLF
 251 ILMELHTISV VFIAAGMERI NTSSEGDIPF CINAKEPIK DTFMALAALS
 301 KVCATLT*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF24 shows 96.4% identity over a 307 aa overlap with an ORF (ORF24a) from strain A of *N. meningitidis*:

	10	20	30	40	50	60
orf24a.pep	MRTAVVLLIMPMMAASSAMMPEMVCAGVSPGTAILISXPTQTAIVASSLSNVSTPASAAA					
orf24	MRTAVVLLIMPMMAASSAMMPEMVCAGVSPGTAILISKTEQTAVMASSLSVSTPASAAA					
	10	20	30	40	50	60
orf24a.pep	IIPSSSXTGINAPLKPPTALEAIMPPFFTASFNSAKAAVVCVPQTLKPISSRMRATESP					
orf24	IIPSSSXTGINAPLKPPTALEAIMPPFFTASFNSAKAAVVCVPQTLKPISSRMRATESP					
	70	80	90	100	110	120
orf24a.pep	TAGVGASDKSRIPNGIFSI EASRPMSSPTRVILKAVFFTTSATSVNVVA SEFSNAFFT					
orf24	TAGVGASDKSRIPNGIFSI EASRPMSSPTRVILKAVFFTTSATSVNVVA SEFSNAFFT					
	130	140	150	160	170	180
orf24a.pep						
orf24						

		130	140	150	160	170	180
5	orf24a.pep	190	200	210	220	230	240
		FGPDTPTLITASASPEFXNAPAIKGLSSXALQNTTILAQPKPSSVISXVRLMVSPASLTA					
	orf24	190	200	210	220	230	240
		FGPDTPTLITASASPEFXNAPAIKGLSSXALQNTTILAQPKPSSVISXVRLMVSPASLTA					
10	orf24a.pep	250	260	270	280	290	300
		SILIPARVLPILMELHTISVVFASGNERXNTSSEGDIPFCTSAEKPPIKDTMPALAALS					
	orf24	250	260	270	280	290	300
		SILIPARVLPILMELHTISVVFASGNERXNTSSEGDIPFCTSAEKPPIKDTMPALAALS					
15	orf24a.pep	KVCATLTXX					
	orf24	KVCATLTXX					

The complete length ORF24a nucleotide sequence <SEQ ID 677> is:

20	1	ATGGCGCAGG	CAGTGGTITT	GCTGTGATC	ATGCCGATGG	CGGCTTCGTC
	51	GGCAATGATG	CCGGAATAGG	TGTGCGCGGG	TGTGTCGCGG	GGAACGGCAA
	101	TCATATCCAA	NCCGACCGAA	CAAAACGGCG	TCATGCTTC	GAGTTTATCC
	151	AACGTGACGA	CGCTGCTTC	GCGCGCGCA	ATCATACCTT	CGTCTCCGGA
25	201	NACGGGGATA	AACGCGCCAC	CAAAACCGCC	AACCGCGCTC	GAAGCCATCA
	251	TGCGCCGCTT	TTTCAAGCCA	TGTTTACGCA	ATGCCAAAGC	TGCTGTTGTG
	301	CGGTGCGTAC	CGCAGACGCT	CAAAACCAT	TCITCAAGAA	TGCGCGCCAC
	351	CGAGTCGCGC	ACGCGAGGGG	TGGTGCACG	CGACAAGTCG	AGARTACCAA
	401	ACGGGATATT	CAGCATTTTT	GAGGCTTCGC	GGCGATGAG	TTGCGCCACG
	451	CGGTAATTTT	TGAAGCGGCT	TTTCTTCACA	ACTTCGGCAA	CTTCGGTCAA
30	501	TGTGCTTGCA	TCCGAATTTT	CAACGCGCG	TTTTAGSACA	CCCGGCGCGG
	551	ATACGCGCAG	ATTAATCACR	GCATCCGCTT	CGCCTAGGCC	GTGACACGCG
	601	CCGCGCATAM	ACGGGTTGTC	TTCGCGCGG	TTGCGACACA	CGACGATTTT
	651	CGCGCAGCGG	AACCTTCTCA	GTGTGATTTT	ANCCGTGCGT	TGTGATGTTT
	701	CGCGCGCCAG	TCTGACCGCG	TCCATATTGA	TACCGCGCGG	CGTACTGCGG
35	751	ATATTGATGG	AGCTGCACAC	GATATCAGTA	GTCTTATCCG	CTTCGGGAAT
	801	GGACGCGATN	AACACCTCGT	CAGAAGCGGA	CATACCTTTT	TGCACCGAGG
	851	CGGAAAGGCC	GCCAATAAAA	GACACGCCGA	TGGCTTTGGC	AGCCTTATTC
	901	AAAGTTTGGC	CCACGCTGAC	GTA		

This encodes a protein having amino acid sequence <SEQ ID 678>:

40	1	MRTAVVLLLI	MPMAASSAMM	PEMVCAGVSP	GTALISXPT	QTAVIASSLS
	51	NVSTPASAAA	IIPSSSXTGI	NAPLKPPTAL	EAIMPPFFTA	SFSNAKAADV
	101	PCVPQTLKPI	SSRMRAATESP	TAGVGASDKS	RIPNGIFISF	EASRPMSSPT
	151	RVILKAVEFT	TSATSVNVVA	SEFSNAAEFT	PGDPTPTLIT	ASASPEPNNA
45	201	PAIXGLSSXA	LQNTTILAQP	KPSSVISXVR	LMVSPASLTA	SILIPARVLP
	251	ILMELHTISV	VFIASGMRX	NTSSEGDIPF	CTSAEKPPIK	DTMPALAALS
	301	KVCATLT*				

It should be noted that this protein includes a stop codon at position 198.

ORF24a and ORF24-1 show 96.4% identity in 307 aa overlap:

		10	20	30	40	50	60
50	orf24a.pep	MRTAVVLLLI	MPMAASSAMM	PEMVCAGVSP	GTALISXPT	QTAVIASSLS	NVSTPASAAA
	orf24-1	MRTAVVLLLI	MPMAASSAMM	PEMVCAGVSP	GTALISXPT	QTAVIASSLS	NVSTPASAAA
		10	20	30	40	50	60
55	orf24a.pep	70	80	90	100	110	120
		IIPSSSXTGINAPLKPPTALEAIMPPFFTASFSNAKAADVPCVPQTLKPISSRMRAATESP					
	orf24-1	70	80	90	100	110	120
		IIPSSSXTGINAPLKPPTALEAIMPPFFTASFSNAKAADVPCVPQTLKPISSRMRAATESP					
60		130	140	150	160	170	180

	orf24a.pep	TAGVGASDKSRIPNGIIFSIFEASRPMSSPTRVILKAVFFTTSATSVNVVASEFNSAAFTT	
	orf24-1	TAGVGASDKSRIPNGIIFSIFEASRPMSSPTRVILKAVFFTTSATSVNVVASEFNSAAFTT	
5		130 140 150 160 170 180	
	orf24a.pep	PGPDTPTLITASASPEPXNAPAIKGLSSXALQNTTILAQPKPSSVIXVRLMVPSPASLTA	
	orf24-1	PGPDTPTLITASASPEPXNAPAIKGLSSXALQNTTILAQPKPSSVIXVRLMVPSPASLTA	
10		190 200 210 220 230 240	
	orf24a.pep	SILIPARVLPIILMELHTISVVFASGMRXNTSSEGDIPFCTSAEKPPKIDTTPMALAALS	
	orf24-1	SILIPARVLPIILMELHTISVVFASGMRXNTSSEGDIPFCTSAEKPPKIDTTPMALAALS	
15		250 260 270 280 290 300	
	orf24a.pep	KVCATLTIX	
	orf24-1	KVCATLTIX	
20			

Homology with a predicted ORF from *N.gonorrhoeae*

ORF24 shows 96.7% identity over a 121 aa overlap with a predicted ORF (ORF24ng) from

25	<i>N.gonorrhoeae</i> :		
	orf24.pep	MRTAVVLLILIMPMAASSAMPMEMVAGVSPCTAIISKPTQTAVMMASSLSVSTPASAAA	60
	orf24ng	MRTAVVLLILIMPMAASSAMPMEMVAGVSPGTAIMSKPTEQTAVMMASSLSVNTPASAAA	60
30	orf24.pep	IIPSSSETGINAPLKPTALEAIMPFFFTASFSNAKAAVVCVPQTLKPIIXSRMRATKSP	120
	orf24ng	IIPSSSETGINAPLKPTALEAIMPFFFTASFSNAKAAVVCVPQTLKPIISSRMRATESP	120
35	orf24.pep	TG	122
	orf24ng	TAGVGASDKSRMPNGIIFSIFEASRPMSSPTRVILKAVFFTTSATSVRLTASEFSSAALT	180

The complete length ORF24ng nucleotide sequence <SEQ ID 679> is:

40	1	ATGCGCACGCG	CGGTGGT	TTT	GCTGTGATC	ATGCGGATGG	CGGCTTCGTC
	51	GGCGATGATG	CGGAAATGG	TGTGCGCGGG	CGTGTGCGCG	GGAAACGGCAA	
	101	TCATGTCGCA	ACCAACGCGAG	CAGACGCGG	TCAGGCGGTC	GAGTTTGTCC	
	151	AGCGTCAACA	CGCTGCGTC	GCGCGCGCA	ATCATACCTT	CGCTTCGGA	
	201	AACGGGATA	AACGCGCGC	TCAAACCGCC	GACCGCGTG	GAAAGCATCA	
	251	TGCCGCGCTT	TTTACGCGCA	TGCTTCAGCA	ATGCCAAAGC	TGCTGTTGTG	
	301	CGGTGCGTAC	CGCAGACGCT	CAAGCCCAT	TCTTCAAGAA	TGCGCGCAC	
45	351	CGAGTCGCGG	ACGCGCGGGG	TGGTGGCCAG	CGACAAATCG	AGAATCGCGA	
	401	ACGGGATATT	CAGCATTTT	GAGGCTTCGC	GACCGATGAG	TTGCGCCACG	
	451	CGSGTGATTT	TGAAAGCGGT	TTTCTTACG	ACTTCGCGCA	CCTCGGTGAG	
	501	GCTGACGCGG	TCCGAATTT	CCAGCGCGGC	TTTGACACG	CCTGGACCGG	
50	551	ATACGCGCAG	ATTAATACCA	GCATTCGCTT	CGCCGAGGCC	GTGGACCGCA	
	601	CCGCGCATAA	ACGGATTGTC	TTCCACCGCG	TTCGAGAAC	CGAAGATTTT	
	651	GGCGCAGCGG	AAACCTTCGG	GTGTGATTT	AGCCGTGCGT	TTGATGGTTT	
	701	CGCCTGCGAG	CTTGACCGCA	TCCATATTGA	TACCGGCACG	CGTGTGCGCG	
	751	ATATTGATGG	AGCTGCACAC	GATATCGGTA	GTTTTCATCG	CTTGGGAAC	
	801	GGACCGGATC	ACACCTGAT	CCGAGGGCGA	CATACCTTTT	CTTGACACG	
55	851	CGGAAGAACC	CCCATATAAG	GACACGCCGA	TGGCTTGGC	TGCTTGTGCG	
	901	AAAGCTGCG	CCACGCTGAC	ATAA			

This encodes a protein having amino acid sequence <SEQ ID 680>:

60	1	MRTAVVLLI	MEMAASSAMM	PEMVAGVSP	GTAIMSKPTE	QTAVMASSLS
	51	SVNTPASAAA	IIPSSSETGI	NAPLKPTAL	EAIMPFFFTA	SFSNAKAAVV
	101	PCVPQTLKPI	SSRMRATESP	TAGVGASDKS	RMPNGIISIF	EASRPMSSPT
	151	RVILKAVFFT	TSATSVRLTA	SEFSSAALT	PGPDPTLIT	ASASPEWNA
	201	PAINGLSSTA	LQNTTILAQP	KPSGVISAVR	LMVSPASLTA	SILIPARVLP

251 ILMELHTISV VFIASGTERI NTSSEGDIPF CTSAEKPPIK DTPMALAALS
 301 KVCATLT*

ORF24ng and ORF24-1 show 96.1% identity in 307 aa overlap:

		10	20	30	40	50	60
5	orf24-1.pep	MRTAVVLLIMPMAASSAMMP	EMVCAGVSPGT	AIISKPTQ	AVMASSLSV	STPASAAA	
	orf24ng	MRTAVVLLIMPMAASSAMMP	EMVCAGVSPGT	AIMSKPTQ	AVMASSLSV	STPASAAA	
		10	20	30	40	50	60
10	orf24-1.pep	IIPSSSETGINAPLKPP	TALEAIMP	PPFTASFS	NAKA	AAVVP	CVPTLKF
	orf24ng	IIPSSSETGINAPLKPP	TALEAIMP	PPFTASFS	NAKA	AAVVP	CVPTLKF
		70	80	90	100	110	120
15	orf24-1.pep	TAGVGASDKSRIPNG	IFSI	FEASRFMSS	PTRVILK	AVFF	TTTSATSV
	orf24ng	TAGVGASDKSRMPNG	IFSI	FEASRFMSS	PTRVILK	AVFF	TTTSATSV
		130	140	150	160	170	180
20	orf24-1.pep	PGPDTPTLITASAS	PEPNAPAIN	GLSSTALQ	NTTILAQ	PKPSG	IVSAVRL
	orf24ng	PGPDTPTLITASAS	PEPNAPAIN	GLSSTALQ	NTTILAQ	PKPSG	IVSAVRL
		190	200	210	220	230	240
25	orf24-1.pep	SILIPARVLP	ILMELHTIS	VVF	FIASGME	INTS	SEGDIP
	orf24ng	SILIPARVLP	ILMELHTIS	VVF	FIASGME	INTS	SEGDIP
		250	260	270	280	290	300
30	orf24-1.pep	ATACGATATC	TTGCGGCTA	TTCCATCGAT	TAG		
	orf24ng	ATACGATATC	TTGCGGCTA	TTCCATCGAT	TAG		
		250	260	270	280	290	300
35	orf24-1.pep	KVCATLT					
	orf24ng	KVCATLT					

Based on this analysis, including the presence of a putative leader sequence (first 18 aa – double-underlined) and putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 81

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 681>:

45 1 ..ACCGACGTGC AAAAAGATT GGTGCGCGAA CAACGCAAGT GGGCGCAGGA
 51 AAAAAATCAGC AACTGCGGAC AAGCGCGCGC GCAGGCAGAC CGGCAGGAAT
 101 ACGCCGAATA CTTCAAGCTG CAATGCGGACA CGCGGATGAC GCGCGAARCG
 151 ATACGATATC TTGCGGCTA TTCCATCGAT TAG

This corresponds to the amino acid sequence <SEQ ID 682; ORF25>:

50 1 ..TDVQKELVGE QRKWAQEKIS NCRQAAQAD ROEYAEYLKL QCDTRMTRER
 51 IQYLRGYSID *

Further work revealed the complete nucleotide sequence <SEQ ID 683>:

1 ATGTATCGGA AACTCATTGC GTCGCCGTTT GCCCTGCTGC TTGCCGCTTG
 51 CGGCAGGGAA GAACCGCCCA AGGCATTGGA ATGCGCCAC ACCTCCGCTGT

5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
95
100

101 TGCAAGGCAT ACGCGGCAAT ATTCAGGAAA CGCTACGCGA GGAAGCGCGT
151 TCCTTTGCGC GCGAGACGCG CAGGCAGATT GTCGATGCGC ACNAAATTAT
201 GCGCGCGCGC TACGGTTTGG CGTTTCTCTT GGAACACGCT TCGGAAACGC
251 AGCAAGGCGC GCGCACGCTC TGATTCGCGC ATTTGAACAT TACCGTGCGC
301 TCTGAAACGC TTGCGCGATG CAAGGCAAC AGCCCCCTGT TGTACGGGGA
351 AACTGCTTTG TCGGATATTG TCGGCGAGAA GACGGCGCGC AATGTCGAGT
401 TTAAGACGCG CGTATTGACG GCAGCCGTCC GCTTCTGCGC CGTCAAGAC
451 GGTACAGCGC CATTTGTGCA CAACACGCTC GGTATGCGCG CGCAACGCT
501 GTCTGCGCGC CTGCTGCCTT ACGCGGTGAA GAGCATCGTG ATGATAGACG
551 GCAAGGCGGT GAAAAAGAA GACGCGGTCA GGATTTTGAG CGGAAAGGCC
601 CGTGAAGAAG AACCGTCCAA ACCCACGCC AGACCATTTT TGGAAACAA
651 TGCGCCGCGC GCGCATGCGG GCGTACCCCA AGCCGCGAGAA GCGCGCGCGC
701 AACCGGAAT CTGCATCTCT GACGACGCGC AGCGTGCGGA TACCGTTACC
751 GTATCAGCGG GCGAAGTGGG AGAGGCGCGC GTACAAAACC AGCGTGCGGA
801 ATCCGAATTT ACCAAACTTT GGGGAGGACT CGATACCGAC GTGCAAAAAG
851 AGTTGGTCCG CGAACACGCG AAGTGGCGCG AGGAAAAAAT CAGCAACTGC
901 CGACAAGCGC CCGCGCAGCG AGACGCGCG GATACGCGG AATACCTCAA
951 GCTGCAATCG GACACGCGGA TGACGCGGGA ACGGATACAG TATCTTCGCG
1001 GCTATTCCAT CGATTAG

20 This corresponds to the amino acid sequence <SEQ ID 684; ORF25-1>:

25
30
35
40
45
50
55
60
65
70
75
80
85
90
95
100

1 MYRKLIALPF ALLLAACGRE EPPKALECAN PAVLQIGRIN IOETLTQEAR
51 SFAREDGRQF VDADKTI AIAA YGLAFSLAHE SETQEGGRTF CIADLNTVP
101 SETLADAKAN SPLLYGETAL SDIVRQKTGG NVEPKDGLVT AAVRFLPVKD
151 GQTAFVNDTV GMAAQTL SAA LLLPYGVKIV MIDGKAVKKE DAVRILSGKA
201 REEESKPEPT EDILEHNAAG GDAGVPQAAE GAPEPILPH DDGERADTVT
251 VSRGEVEEAR VQNQRAESEI TKLWGLDLTD YKELVGEQR KWAQEKISNC
301 RQAAQADRQ EYAEYLKLCQ DTRMTRERIQ YLRGYSID*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

30 ORF25 shows 98.3% identity over a 60aa overlap with an ORF (ORF25a) from strain A of *N.*

meningitidis:

35
40
45
50
55
60
65
70
75
80
85
90
95
100

orf25.pep
250 260 270 280 290 300
VTVSRGEVEEARVQNQRAESEITKLWGLDLTDVQKELVGEQRKWAQEKISNCRQAAQAD
TDVQKELVGEQRKWAQEKISNCRQAAQAD
orf25a
250 260 270 280 290 300
RQEYAEYLKLCQDTRMTRERIQYLRGYSIDX
RQEYAEYLKLCQDTRMTRERIQYLRGYSIDX
orf25a
310 320 330

The complete length ORF25a nucleotide sequence <SEQ ID 685> is:

45
50
55
60
65
70
75
80
85
90
95
100

1 ATGTATCGGA AACTCATGTC GCTGCGGTTT GCGCTGCTGC TTGCGCGCTTG
51 CGGACGGGAA GAACCGCGCA AGGCATTTGGA ATTCGCGGCT CCGCGCGGT
101 TGCAAGGACAT ACGGCGCAAT ATTCAGGAAA CCGTACGCGA GGAAGCGGCT
151 TCCTTTGCGCG GCGAGACGNG CANGCAGATT GTCGATGCGC ACNAAATTAT
201 GCGCGCGCGC TANGNTNNGN NGNTNTCTTT GGAACACGCT TCGGAAACGC
251 AGGAAGCGCG GCGCACGCTC TGNTGCGCG ATTTGAACAT TACCGTGCGC
301 TCTGAAACGC TTGCGGATGTC CAAGGCAAC AGCCCCGTGC GTACGGGGA
351 AACCGCTTTG TCGGATATTG TCGGCGAGAA GACGGCGCGC AATGTCGAGT
401 TTAAGACGCG CGTATTGACG GCAGCCGTCC GCTTCTACCG CGTCAAGAC
451 GGTACAGANG CATTTGTGCA CAACACGCTC GGTATGCGCG CGCAACGCT
501 GTCTGCGCGC TTGCTGCCTT ACGCGGTGAA GAGCATCGTG ATGATAGACG
551 GCAAGGCGGT AAAAAAGAA GACGCGGTCA GGATTTTGAG CNGANAAGCC
601 CGTGAANAAG AACCGTCCAA ANCCNNGCCC AGACCATTTT TGGAAACATA
651 TGCGCGCGGA GGGGATGCGC ACGTACCCCA AGCCGGAGAA GACGCGCGCG
701 AACCGGAAT CTGCTATCTT GAGACGCGCG ACGCTGCGCA TACCGTTACC
751 GTATCGCGG GCGAAGTGGG AGAGGCGCGN GTACAAACCC ACGCTGCGGA
801 ATCCGAATTT ACCAAACTTT GGGGAGGACT CGATACCGAC GTGCAAAAAG

```

851 AGTTGGTCGG CGAANAACGC AAGTGGGCGC AGGAAAAAAT CAGCAACTGC
901 CGACAAGCCG CCGCGCAGCG AGACCGGCAG GAATACGCCG AATACCTCAA
951 GCTGCMATGC GACACGCGGA TGACGCGCGA ACGGATACAG TATCTTCGCG
1001 GCTATTCCAT CGATTAG

```

5 This encodes a protein having amino acid sequence <SEQ ID 686>:

```

1 MYRKLIALPF ALLLAACGRE EPPKALECAN PAVLQXIRXN IQETLTQEAR
51 SFAREDXXQF VDADKI IAAA XXXXSLEHA SETQEGGRTF CXADLNIIVP
101 SETLADAKAN SPILLYGETAL SDIVRQKTTGG NVEFKDGVLT AAVRFLPVKD
151 GQXAFVDNTV GMAAQTLSAA LLYPGVKSIV MIDGKAVKKE DAVRIXSXXA
201 REXEPSKXFP EDILEHNAAG GDADVQAGE DAPEPEILHP DDGERADTVT
251 VSRGEVEEAR VQNQRAESEI TKLWGGLD TDVQKELVGE XR KWAQEKISNC
301 RQAAQADRQ EYAEYLKQC DTRMTRERIQ YLRGYSID*

```

ORF25a and ORF25-1 show 93.5% identity in 338 aa overlap:

```

15 orf25a.pep      10      20      30      40      50      60
      MYRKLIALPFALLLAACGRE EPPKALECANPAVLQXIRXNIQETLTQEARSFAREDXXQF
      |||
orf25-1      MYRKLIALPFALLLAACGRE EPPKALECANPAVLQGI RGNIQETLTQEARSFARDGROF
      |||
      10      20      30      40      50      60

20 orf25a.pep      70      80      90      100     110     120
      VDADKIIAAAXXXXSLEHASETQEGGRTFCXADLNIIVPSETLADAKANSPLLYGETAL
      |||
orf25-1      VDADKIIAAAYGLAFSLEHASETQEGGRTFCIADLNIIVPSETLADAKANSPLLYGETAL
      |||
      70      80      90      100     110     120

25 orf25a.pep      130     140     150     160     170     180
      SDIVRQKTGGNVEFKDGVLTAAVRFLPVKDGGQXAFVDNTVGM AQTLSAALLPYGVKSIV
      |||
orf25-1      SDIVRQKTGGNVEFKDGVLTAAVRFLPVKDGGTAFVDNTVGM AQTLSAALLPYGVKSIV
      |||
      130     140     150     160     170     180

30 orf25a.pep      190     200     210     220     230     240
      MIDGKAVKKE DAVRIXSXXAREXEPSKXFPEDILEHNAAGDADVQAGEDAPEPEILHP
      |||
orf25-1      MIDGKAVKKE DAVRILSGKAREEESKPTPEDILEHNAAGDAGVQA AEGAPEPEILHP
      |||
      190     200     210     220     230     240

35 orf25a.pep      250     260     270     280     290     300
      DDGERADTVTVSRGEVEEARVQNQRAESEITKLWGGLD TDVQKELVGE XRKWAQEKISNC
      |||
orf25-1      DDGERADTVTVSRGEVEEARVQNQRAESEITKLWGGLD TDVQKELVGE QRKWAQEKISNC
      |||
      250     260     270     280     290     300

40 orf25a.pep      310     320     330     339
      RQAAQADRQ EYAEYLKQC DTRMTRERIQYLRGYSIDX
      |||
orf25-1      RQAAQADRQ EYAEYLKQC DTRMTRERIQYLRGYSIDX
      |||
      310     320     330

```

50 Homology with a predicted ORF from *N.gonorrhoeae*

ORF25 shows 100% identity over a 60aa overlap with a predicted ORF (ORF25ng) from *N.gonorrhoeae*:

```

55 orf25.pep      TDVQKELVGEQRKWAQEKISNCRQAAAQAD      30
      |||
orf25ng      VTVSRGEVEEARVQNQRAESEITKLWGGLD TDVQKELVGEQRKWAQEKISNCRQAAAQAD      308
      |||
orf25.pep      RQ EYAEYLKQC DTRMTRERIQYLRGYSID      60
      |||
orf25ng      RQ EYAEYLKQC DTRMTRERIQYLRGYSID      338

```

60 The complete length ORF25ng nucleotide sequence <SEQ ID 687> is:

1 ATGTATCGGA AACTCATTGC GGTGCCGTTT GCCCTGCTGC TTGCAGCGTG
 51 CGCGAGGGA GAACCGCCCA AGCGGTGGGA ATGCGCCAAC CCCGCCCGTG
 101 TGCAGGACAT ACGCGGCAGT ATTTCAGGAA GCCTCAACGA GGAAGCGCGT
 151 TCTTTCGCGC GCGAAGACGG CAGGCAGTTT GTCGATCCGG ACAAAATATAT
 201 CGCGCGCGCC TACGGTTTGG CGTTTTCTTT GGAAACAGCT TCGGAAACGC
 251 AGGAAGCGGG GCGCAGCTTC TGTATCGCGG ATTTGAACAT TACCGTGCGG
 301 TCTGAAACGG TTGCGGATGC CGAGGCAAAAC AGCCCCCTGC TGTATGGGGA
 351 AACGCTCTTG CGACACATCG TGCAGCAGAA GACGGCGCGC AATGTCGAGT
 401 TTAAGACGG CGTATTGAGC GCAGCCGTCC GCTTCTGCCC CGCCAAAGAC
 451 GCTCGGACGG CATTTATCGA CAACACGCTC GGATGCGGGA CGCAAAACGT
 501 GTCTCGCGCG TATGTCGCTT ACGGCGTGA GAGCATCGTG ATGATAGACG
 551 GCAAGCGCGT GACAAAAGAA GACGCGGTCA GGGTTTTGAG CGGCAAGCGT
 601 CGTGAAGAAG AACCGTCCAA ACCCAACCCC GAAGACATTT TGAACACCAA
 651 TGCGCGCGGG GCGGATGCGG GCGTACCCCA AGCCCGCAGAA GCGCCACCGG
 701 AACCGCGAAT CCTGCATCCC GAGCAGCTGC AGCTCCCGA TACCGTTACC
 751 GTATCAGCGG CGGAGTCCGA AGAGCGCGCG TACACAAAC AACCTGCGGA
 801 ATCCGAAATT ACCAACTTT GGGGAGACT CGATACCCAG GTGCAAAAG
 851 AGTTGGTCGG CGAACACGCG AAGTGGCGCG AGGAAAAAT CAGcaactgc
 901 CjACAGCCG CCGCGCAGCG AGACCGGCG GATACGCGG AATACCTCAA
 951 GCTCCAATGC GACACGCGGA TGACGCGCGA ACggaTACAG TATCTTCGCG
 1001 GCTATTCCAT CGATTAG

This encodes a protein having amino acid sequence <SEQ ID 688>:

1 MYRKLIALPF ALLLAACGRE EPKALECAN PAVLQDIRGS IQETLTQEAR
 51 SFAREDRGRF VDADKIIAAY YGLAFSLEHA SETQEGGRTF CIADLNTVP
 101 SETLADAEAN SPLLYGETSL ADIVQKGTGG NVEFKDGLVT AAVRFLPAKD
 151 ARTAFIDNTV GMATQTLsAA LLYPVGVKSIV MIDGKAVTKE DAVRVLsGKA
 201 REEESKPTF EDILEHNAAG GDAGVEQAAE GAPEPEILHP DDVERADTVT
 251 VSRGEVEEAR VQNRAESEI TKLWGLDLDV VQKELVGEQR KWAQEKISNC
 301 RQAAQADRDQ EYAEYLKLCQ DTRMTRERIQ YLRGSYID*

ORF25ng and ORF25-1 show 95.9% identity in 338 aa overlap:

		10	20	30	40	50	60
orf25-1.pep	MYRKLIALPF	ALLLAACGRE	EPKALECAN	PAVLQDIRGS	IQETLTQEAR	SFAREDRGRF	
orf25ng	MYRKLIALPF	ALLLAACGRE	EPKALECAN	PAVLQDIRGS	IQETLTQEAR	SFAREDRGRF	
		70	80	90	100	110	120
orf25-1.pep	VDADKIIAAY	YGLAFSLEHA	SETQEGGRTF	CIADLNTVP	SETLADAEAN	SPLLYGETAL	
orf25ng	VDADKIIAAY	YGLAFSLEHA	SETQEGGRTF	CIADLNTVP	SETLADAEAN	SPLLYGETSL	
		130	140	150	160	170	180
orf25-1.pep	SDIVRQKTG	GNVEFKDGLVT	AAVRFLPAKD	ARTAFIDNTV	GMATQTLsAA	LLYPGVKSI	
orf25ng	ADIVQKGTG	GNVEFKDGLVT	AAVRFLPAKD	ARTAFIDNTV	GMATQTLsAA	LLYPGVKSI	
		190	200	210	220	230	240
orf25-1.pep	MIDGKAVTKE	DAVRVLsGKA	REEESKPTF	EDILEHNAAG	GDAGVEQAAE	GAPEPEILHP	
orf25ng	MIDGKAVTKE	DAVRVLsGKA	REEESKPTF	EDILEHNAAG	GDAGVEQAAE	GAPEPEILHP	
		250	260	270	280	290	300
orf25-1.pep	DDGERADTVT	VSRGEVEEAR	VQNRAESEI	TKLWGLDLDV	VQKELVGEQR	KWAQEKISNC	
orf25ng	DDVERADTVT	VSRGEVEEAR	VQNRAESEI	TKLWGLDLDV	VQKELVGEQR	KWAQEKISNC	
		310	320	330	339		
orf25-1.pep	RQAAQADRDQ	EYAEYLKLCQ	DTRMTRERIQ	YLRGSYID			
orf25ng	RQAAQADRDQ	EYAEYLKLCQ	DTRMTRERIQ	YLRGSYID			

Based on this analysis, including the presence of a predicted prokaryotic membrane lipoprotein lipid attachment site (underlined) in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF25-1 (37kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 16A shows the results of affinity purification of the GST-fusion protein, and Figure 16B shows the results of expression of the His-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for Western blot (Figure 16C), ELISA (positive result), and FACS analysis (Figure 16D). These experiments confirm that ORF25-1 is a surface-exposed protein, and that it is a useful immunogen.

Figure 16E shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF25-1.

Example 82

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 689>

```

1   ATCGAGCTGA   TCGACTATTC   ACATTCAATT   TTCTCGGTTG   TGCCACCCCTT
51  TTTTGGCACTG   GCACCTTGCCG   TCATTACCCG   CCGCGTACTG   CTGCTTTTAG
101 GCATCGGTAT   TCTGwysGCG   GTTGCCITTT   TGGTCGGCGG   CAACCCCGTC
151 GACGGTCTGA   CACACCTGAA   AGACATGGTC   TCGGCGTTGG   CTTGGTCAGA
201 CGsyGATTGG   TCGCTGGGCA   AACCAAAAT   CTTGGTTTTC   CKGATACCTT
251 TGGGTATTTT   TACTTCCCTG   CTGACCTACT   CCGGCAGCAA   T.....

                               //
851 .....          .....          .....          .....          AC   TTCTCGGTGA
901 TTGGCGGCGCA   CTTGCGGCGT   CTTTGGCCGT   GTTCTCTGCA   CGCTCGGCAC
951 GATTAAAACC   GCGCACTATC   CCAAAGCCGT   TTGGCAGGGT   GCGAAATCTA
1001 TGTTCCGGCG   AATCGCCATT   TTAATCCTCG   CTTGGCTCAT   CAGTACGGTT
1051 GTGGCGGAAA   TGCACACCGG   CGATTACCTC   TCCACACTGG   TTGGCGGCRA
1101 CATCATCC   GGTCTCCG   CCGTCACT   CTTCTCGCTC   GCGACGCTGA
1151 TGGCGTTTC   CACAGGCACA   AGCTGGGGGA   CTTTCGGCAT   TATGTCGCGC
1201 ATTGCGCGCG   CCAATGCGGT   CAAAGTCGAA   CCGCGCGCTGA   TTATCCCGTG
1251 TATGTCGCGCA   GTAATGCGGG   GGGCGGTATG   CCGCGACACC   TGCTCGGCCA
1301 TTTCCGACAC   GACCATCCTG   TCGTCCACCG   GCGCGCGCTG   CAACACATC
1351 GACCAGCTTA   CCTCGCAACT   GCCTTACGCC   TTAACCGITT   CCGCCGCGCG
1401 CGCATCGGGC   TACCTCGCAT   TGGGCTCTGAC   AAAATCCGCG   CTGTTCGGCT
1451 TTGGCAGCAC   AGGCATTGTA   TTGGCGGTGC   TGATTTTCT   GTTGAAAGAT
1501 AAAAAA..

```

This corresponds to the amino acid sequence <SEQ ID 690; ORF26>:

```

1   MQLIDYSHSF   FSVVPPFLAL   ALAVITRRVL   LSLGIGILXX   VAFVLGNGNPV
51  DGLTHLKMV   VGLAWSDXDW   SLGPKILVF   XILLGIFTSL   LTYSGSN...

                               //
251 .....          .....          .....          .....          TSLV
301 PGTCGVPAV   VLCTLGTKT   ADYFKAWQG   AKSMFGATAI   LLLAWLSTV
351 VGEHMTGDYL   STLVAENIHP   GFLEVILLFL   ASVMFAATGT   SWGTFGIMLP
401 IAAAMAVKVE   PALIIPCMSA   VMAGAVCGDH   CSPISDTTIL   SSTGARNHI
451 DHVTSQLEPYA   LTVAAAAASG   YLALGLTKSA   LLGFGTGTIV   LAVLIFLLKD
501 KK..

```

Further work revealed the complete nucleotide sequence <SEQ ID 691>:

```

1 ATGCAGCTGA TCGACTATTC ACATTCAATT TTCTCGGTTG TGCCACCCTT
51 TTTGGCACTG GCACITTCGCC TCATTACCCG CCGCGTACTG CTGCTTTAG
101 GCATCGCTAT TCTGGTGGCG GTTGGCTTTT TGGTCGGCGG CAACCCGCTC
151 GACGGTCTGA CACACCTGAA AGACATGGTC GTGCGCTTGG CTTGGTCAGA
5 201 CGCGGATTGG TCGCTGGGCA AACCAAAAT CTTGGTTTTC CTGATACTTT
251 TGGGTATTTT TACTTCCTGT CTGACTACT CCGGCAGCAA TCAGGCGGTT
301 GCGCACTGGG CAAAACGGCA CATTAAAAAC CGCGCGCGCG CGAAAATGCT
351 GACCGCTGCG CTCGTGTTGC TAACCTTTAT CGACGACTAT TTCCACAGCTC
401 TCGCGCTCGG TCGGATTCGC CGCCCGGTTA CCAGCAAGTT TAAAGTTTTC
10 451 CGCACCAAAAC TCGCTACATC CCTCGACTCC ACTGCGCGCTC CTATGTGGGT
501 GCTGATGCC GTTCAAGCT GGGGCGCGTC GATTATCGCC AGCGTTCCGC
551 GACTGCTCGT TACCTACAAA ATCACCGAAT ACACGCGCAT GGGGACGGTT
601 GTGCGCATGA GCGTGATGAA CTATTACGCA CTGTTTGCCG TGATTATGCT
15 651 GTTCGTCGTC GCATGGTTTT CCTTCGACAT CGGCTCGATG GCACGTTTGC
701 AACCAAGCGCG GTTGAACGAA CGCCACGATG AACTCGCTGT TTCAGACGCT
751 ACCAAAGGTC GTGTTTACGC GCGCCAGATG CCGCTTTCGG CCTTAATCGC
801 CTCACGCTTT TCGCGCATGA TCTACACCGG CCGCAGGCA AGCGAAACCT
851 TCGACATTTT GGGGCGATT TGAACACGCG ACCTAACAC CTGCTTGCTA
901 TTGCGCGGCA CTTCGCGCGT CCTTGCCTGC GTTCTCGCA CGCTCGGCAC
20 951 GATTAAAAAC CGCGACTATC CCAAGCCGT TTGCGAGGGT GCGAAATCTA
1001 TGTTCCGCGC AATGCCATT TTAATCCTCG CTTGGCTCAT CAGTACGGTT
1051 CTGCGCGGAA TGCACACCGG CGATTACCTC TCCACACTGG TTGCGGCGAA
1101 CATCCATCCC GGCTTCCTGC CCGTCATCCT CTTCCTGCTC GCCAGCGTGA
25 1151 TGGCGCTTGC CACAGGCACA AGCTGGGGGA CGTTCGGCAT TATGCTGCGC
1201 ATTGCGCGCG CCATGGCGGT CAAAGTCGAA CCGCGCTGTA TTATCCCGTG
1251 TATGTCGCGA GTAATGCGCG GGGCGGTATG CGCGCACAC TGCTCGGCCA
1301 TTTCCGACAC GACCATCCTG TCGTCCACCG GCGCGCGCTG CAACACACATC
1351 GACCACGTTA CCTCGCAACT GCCTTACGCG TTAACCGTTG CCGCGCGCGC
30 1401 CGCATCGGCG TACCTCGCAT TGGGTCGAC AAAATCCGCG CTGTTGGGCT
1451 TTGGCAGCAC AGGCATTGTA TTGGCGGTG GTTATTTTCT GTTGAAAGAT
1501 AAAAAACGCG CCAACGCGTG A

```

This corresponds to the amino acid sequence <SEQ ID 692; ORF26-1>:

```

1 MQLIDYSHSF FSVVPPFLAL ALAVITRRVL LSLGILVGV VAFVLGGNVP
51 DGLTHLKDML VGLAWSDDGW SLGKPKILVF LILGLITSL LTYSGSNQFV
35 101 ADWAKRHKN RRGAKMLTAC LVEVPTDDY FHSILVGAIA REVTDKFRAS
151 RTKLAYILDS TAAPMCVIMP VSSWGASIIA TLASGLVTVY ITEYTPMGTF
201 VAMSLMNYIA LEALIMVFVV AWFSDIGSM ARFEOAALNE AHDETAVIDA
251 TKGRLVYALII PVLALIASTV SAMITYGAQA SETFSLGAF ENTDVNTSLV
40 351 FGGTCGVLA VLTCLGTIKT ADYKAVWQG AKSMFGAIAI LILAWLISTV
351 VGEHMTGDYL STLVAAGNIHP GFLPVILFLL ASVMAFATGT SWGTFGIMLP
401 TAAAMAVKVE PALIIECMSA VMAGAVCGDH CPSIDTTIL SSTGARNHI
451 DHVTSQLPYA LTVAAAAASG YLAGLTKSA LLEGTTGIV LAVLIFLKD
501 KKRANA*

```

Computer analysis of this amino acid sequence gave the following results:

45 Homology with the hypothetical transmembrane protein HI1586 of *H. influenzae* (accession number P44263)

ORF26 and HI1586 show 53% and 49% amino acid identity in 97 and 221 aa overlap at the N-terminus and C-terminus, respectively:

```

Orf26 1 MQLIDYSHSFSSVPPFLALALAVITRRVXXXXXXXXXXVAFVLGGNVPDGLTHLKDMLV 60
+*+ L T + + TSLV GG C + L + + + +Y ++ G KSM G
50 HI1586 14 MELIDFSSSVSVIPVALLAIILAIATRRVLVSLSAGIIISGLMLSDWQIGSAFNYLVKNV 73

Orf26 61 VGLAWSDXDWSLGPILVFEXILLGIFTSLLTYSGSN 97
V L ++D + + I++ +LLG + T+LLT SGSN
55 HI1586 74 VSLVYADGEIN-SNMNIYVFLLLGLVLTALLTVSGSN 109

//

Orf26 86 IFTSLTYSGS--NTSLVFGGTCGVFAVVLCTL--GTIKTADYPKAVWQGAQSMFGXXXX 141
+*+ L T + + TSLV GG C + L + + + +Y ++ G KSM G
60 HI1586 299 VFSVLGTGFENTVVGTSVLVGGFCSIIISTLLIILDRQVSVPFYVRSWIGIKSMSGAIAI 358

```


Orf26 142 XXXXXXXSTVVGEMHTGDYLSLTVAGNIHPGFLEVLVLLASVMAFATGTSWGTGFMGLP 201
 + +VG+M TG YLS+LV+GNI FLEVILF+L + MAF+TGTSWGTGFMGLP
 HI1586 359 LFFAWTINKIVGDMQTGKYLSSLVSGNIFMQFLPVILFVLGAAMAFSTGTSWGTGFMGLP 418

5 Orf26 202 IAAAMAVKVEPALIIPCMSAVMAGAVCGDHCSPISDITILSSTGARC�NHIDHVT SQXXXX 261
 IAAAMA P L++PC+SAVMAGAVCGDHCSP+SDTITILSSTGA+CNHIDHVT+Q
 HI1586 419 IAAAMAANAPELILPCLSAVMAGAVCGDHCSPVSDTITILSSTGARC�NHIDHVTQLPEYA 478

10 Orf26 262 XXXXXXXXXXXXXXXXXXXXSALLGFGTTGIVLAVLIFLLKDK 302
 S L GF T + L V+IF +K +
 HI1586 479 ATVATATSIGIVVGFYTSGLAGFAATAVSLVILFVAKKR 519

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF26 shows 58.2% identity over a 502aa overlap with an ORF (ORF26a) from strain A of *N.*

meningitidis:

10 20 30 40 50 60
 orf26.pep MQLIDYSHSFFSVVPPFLALALAVITRRVLLSLGIGILXVAFVLVGGNEPVDGLTHLKDMV
 |||||
 orf26a MQLIDYSHSFFSVVPPFLALALAVITRRVLLSLGIGILVGVAFVLVGGNEPVDGLTHLKDMV
 |||||

70 80 90 99
 orf26.pep VGLAWSDXDWSLKGPKILVFXILLGIFTSLTYSGSNXX-----
 |||||
 25 orf26a VGLAWSGDGWSLKGPKXILVFLILLGIFTSLTYSGSNQAFADWAKRHKNRRGAKMLTAC
 |||||

130 140 150 160 170 180
 orf26.pep -----
 orf26a LVFVTFIDDYFHS LAVGAXAREVTDKFKVSRAKLAYILDSTAAPMCVLMFVSSWGASIIA
 |||||

190 200 210 220 230 240
 orf26.pep -----
 orf26a TLAGLLVYTKITEYTPMGTFFVAMSLMNYALFALIMVFFVAVFSDIGSMARFEQAALNE
 |||||

250 260 270 280 290 300
 orf26.pep -----TSILV
 |||||
 orf26a AHDETAVSDGSGWRVYALIPVLALIASTVSAMITYGAQASETFSLGAFTNDVNTSLV
 |||||

310 320 330 340 350 360
 orf26.pep FGGTCGVFAVVLCTGLTIKTADYPKAVWQGAQSMFGAIAILILAWLISTVVGEMHTGDYL
 |||||
 orf26a FGGTCGVFAVVLCTGLTIKTADYPKAVWQGAQSMFGAIAILILAWLISTVVGEMHTGDYL
 |||||

370 380 390 400 410 420
 orf26.pep STLAVAGNIHPGFLEVLVLLASVMAFATGTSWGTGFMGLPFAAAMAVKVEPALIIPCMSA
 |||||
 55 orf26a STLAVAGNIHPGFLEVLVLLASVMAFATGTSWGTGFMGLPFAAAMAVKVEPALIIPCMSA
 |||||

430 440 450 460 470 480
 orf26.pep VMAGAVCGDHCSPISDITILSSTGARC�NHIDHVTSQLPYALTVAANAASGYLALGLTKSA
 |||||
 orf26a VMAGAVCGDHCSPISDITILSSTGARC�NHIDHVTSQLPYALTVAANAASGYLALGLTKSA
 |||||

300 310
 orf26.pep LLGFGTTGIVLAVLIFLLKDKK
 |||||

orf26a LLGFGXTGIVLAVLIFLLKDKKRRANX
490 500

The complete length ORF26a nucleotide sequence <SEQ ID 693> is:

```

1   ATGCAGCTGA TCGACTATT ACATTCAATT TTCTCGGTTG TGCCACCCCTT
5   1   TTGTGGCACTG GCATCTGGCG TCATTACCCG CCGCGTACTG CTGTCTTTAG
101  GCATCGGTAT TCTGGTGGCG GTTGCTTTT TTGTTCGGCG CAACCCCGTC
151  GACGGCTCTGA CACACCTGAA AGACATGGTC GTCCGCTTGG CTTGGTCAGA
201  CGCGGATTGG TCGCTGGGCA AACCAAAANT CTTGTTTTC CTGATACTTT
251  TGGGTATTTT TACTTCCCTG CTGACCTACT CCGGCAGCAA TCAGGCGTTT
10  301  GCGGACTGGG CAAAACGGCA CATTAAAAAC CGGCGCGCG CGAAAAATGCT
351  GACCGCTCTG CTCGTGTTTC TAACCTTTAT CGACACTAT TTCCACAGTC
401  TCGCGCTCGG TCGGNTTGCC GCGCCGCTTA CGGACAAGTT TAAAGTTTCC
451  CGCGCCAAAC TCGCCTACAT COTCGACTCC ACTGCGCGGC CTATGTGGCT
15  501  GCTGATGCCG GTTTCAGACT GGGCGCGGTC GATTATCGCC ACGCTTGGCG
551  GACTGCTCGT TACCTACAAA ATCACCAGAT ACACGCCGAT GGGGACGTTT
601  GTTCGCGATGA GCCTGATGAA CTATTACGCA CTGTTTGGCC TGATTATGCT
651  GTTCGCTCGT CGATGTTCTT CCTTCGACAT CGCTCTGATG CACGCTTTCG
701  AACAGCGCGC GTTGAACGAA GCCACAGAT AACTCGGCTT TCCAGACGCG
751  AGCTGCGGCA GGGTTTACGC ATTGATTATT CCGCTTTTGG CCTTAATCGC
20  801  CTCACGCGTT TCCGCCATGA TCTACACCGG TGCACAGGCA AGCGAAACCT
851  TCAGCATTTT GGGTGCATTT GAAAATACGG ACOTGAACAC TTCGCTGGTA
901  TTCGCGGCGCA CTTGCGGCGT GCTTGCCGTC GTCTCTGCA CGCTCGGCAC
951  GATTAAATCT GCCGATTATC CAAAAGCGGT TTGGCAGGTT GCGAATCCCA
25  1001 TGTTTCGGCG AATCGCCATT TTAATCCTTG CTTGCGTCA CTAGTACGGTT
1051 GTCCGCGGAAA TGCACACAGG CGACTACCTC TCCACGCTGG TTGCGGGCAA
1101 CATCCATCCC GCGTTCTCTG CCGTCACTCT TTCTCGTCTC GCCAGCGTGA
1151 TGGCGTTTTC CACAGGCACA AGCTGGGGGA CGTTGCGCAT CATGCTGGCG
1201 ATTGCGCGCG CGATGCGGCT CAAAGTCGAT CCCTCACTGA TTATCCCGTG
30  1251 TATGTCGCGC GTGATGGCGG GGGCGGTATG CGGCGACCAT TGCTCGGCCA
1301 TTTCCGACAC GACCATCCTG TCGTCCACCG CGCGCGGCTG CAACCCATC
1351 GACACGCTTA CNTCGCAACT GCCTTACGCT TTAACCGTTG CCGCCCGCGC
1401 CGCATCGGGA TACCTCGCAT TGGGCTCGAC AAAATCCGCG CTGTTGGGTT
1451 TTGGCANGAC AGGCATTGTA TTGGCGGTGC TGATTTTCTT GTTGAAAGAT
1501 AAAAAACGCG CCAACGCGTG A

```

This encodes a protein having amino acid sequence <SEQ ID 694>:

```

1   MQLIDYSHSF FSVVPPFLAL ALAVITRRVL LSLGIGILV VAFVLGGNPV
5   1   DGLTHLKMV VGLAWSGDGW SLGKPKXILV LILLGIFTS LITYSGSNQAF
101  ADWAKRHHK RRGAKMLTAC LVFVTFIDDY FHS LAVGAXA RPVTDKFKVS
151  RAKLAYILDS TAAPMCVILMP VSSWGASIIA TLAGLLVTK ITEYTPMGTF
201  VAMSLMNYA LFAALIMVTV AMWFSFDIGSM ARFEQAALNE AHDETAVIDS
251  SWGRVYALII PVLALIASTV SAMIYTGQA SETFSILGAF ENTDVNTSLV
301  FGGTCGVLA VCLTLGTIKI ADYFKAVWQG AKSMFGAIAI LILAWLISTV
351  VGEHMTGDYL STLVAAGNHP GLXVILFLL ASVMAFATGT SWGTFGIMLE
401  TAAAMAVKVD PSLLIPCMSA VMAGAVCGDH CSPISDTTLL SSTGCRNHH
45  451  DHVTSQLPYA LTVAAAAASG YLALGLTRSA LLGFGXTGIV LAVLIFLLKD
501  KKRANA*

```

ORF26a and ORF26-1 show 97.8% identity in 506 aa overlap:

```

10      20      30      40      50      60
orf26a.pep MQLIDYSHSF FSVVPPFLALALAVITRRVLLSLGIGILVGVAFVLGGNPVDGLTHLKMV
10      20      30      40      50      60
orf26-1    MQLIDYSHSF FSVVPPFLALALAVITRRVLLSLGIGILVGVAFVLGGNPVDGLTHLKMV

70      80      90      100     110     120
orf26a.pep VGLAWSGDWLSLGPVKLVFLILLGIFTSLITYSGSNQAFADWAKRHHKNNRRGAKMLTAC
70      80      90      100     110     120
orf26-1    VGLAWSGDWLSLGPVKLVFLILLGIFTSLITYSGSNQAFADWAKRHHKNNRRGAKMLTAC

130     140     150     160     170     180
orf26a.pep LVFVTFIDDYFHS LAVGAXARPVTDKFKVSRKALAYILDS TAAPMCVILMPVSSWGASIIA
130     140     150     160     170     180
orf26-1    LVFVTFIDDYFHS LAVGALARPVTDKFKVSRKALAYILDS TAAPMCVILMPVSSWGASIIA

```

		130	140	150	160	170	180
5	orf26a.pep	190	200	210	220	230	240
	orf26-1	190	200	210	220	230	240
10	orf26a.pep	250	260	270	280	290	300
	orf26-1	250	260	270	280	290	300
15	orf26a.pep	310	320	330	340	350	360
	orf26-1	310	320	330	340	350	360
20	orf26a.pep	370	380	390	400	410	420
	orf26-1	370	380	390	400	410	420
25	orf26a.pep	430	440	450	460	470	480
	orf26-1	430	440	450	460	470	480
30	orf26a.pep	490	500				
	orf26-1	490	500				

Homology with a predicted ORF from *N.gonorrhoeae*

- 40 ORF26 shows 94.8% and 99% identity in 97 and 206 aa overlap at the N-terminus and C-terminus, respectively, with a predicted ORF (ORF26ng) from *N. gonorrhoeae*:

	orf26.pep	MLIDYSHSFSSVVPFFLALALAVITRRVLLSLGIGILXVAFLVGGNPVDGLTHLKMV	60
	orf26ng	MLIDYSHSFSSVVPFFLALALAVITRRVLLSLGIGILVGVAFVGGNPVDGLTHLKMV	60
45	orf26.pep	VGLAWSDXDWSLGGPKILVFLILLGIFTSLITYSGN	97
	orf26ng	VGLAWADGWSLGGPKILVFLILLGIFTSLITYSGNQAFADWAKRHIKNRCGAKMLTAC	120
50		//	
	orf26.pep	TSLVFGGTCGVFAVVLCTLGITKTADYPKA	326
	orf26ng	ASTVSAMIYTGAAQSETFSILGAFENTDVNTSLVFGGTCGVFAVVLCTFGITKTADYPKA	326
55	orf26.pep	VWQAKSMFGAIAILLAWLISTVVGEMHTGDYLSLTVAGNIHGFLEPVLIFLLASVMAF	386
	orf26ng	VWQAKSMFGAIAILLAWLISTVVGEMHTGDYLSLTVAGNIHGFLEPVLIFLLASVMAF	386
60	orf26.pep	ATGTSWGTFGIMLPAAAMAVKVEPALIIPCMASVMAGAVCGDHCSISDITILSSTGAR	446
	orf26ng	ATGTSWGTFGIMLPAAAMAVKVEPALIIPCMASVMAGAVCGDHCSISDITILSSTGAR	446
65	orf26.pep	CNHDHVTSQLPYALTVAASASGYLALGLTKSALLGFGTTGIVLAVLIFLLKDKK	502
	orf26ng	CNHDHVTSQLPYALTVAASASGYLALGLTKSALLGFGTTGIVLAVLIFLLKDKKRAVD	506

The complete length ORF26ng nucleotide sequence <SEQ ID 695> is:

```

1  ATGCAGCTGA  TTGACTATT  ACATTCAATT  TTCTCGGTTG  TGGCACCCCTT
51  TTTGGCACTG  GCACCTGCGG  TCATTACCCG  CCGCGTACTG  CTGCTTTTAG
101  GCATCGGTAT  TTTGGTGGCG  GTTGCCCTTT  TGGTCGGCGG  CAACCCCGCTC
151  GACGGTCTGA  CACACCTGAA  AGACATGTGT  CTGCGCTTGG  CTGGCGAGA
201  CGGCGATTGG  TCGCTGGGCA  AAAAAAATAT  CTGCGCTTGG  CTGTAATCTT
251  TGGGCAATTT  CACTTCACCT  CTGACCTACT  CCGCGAGCAA  TCAGGCGTTT
301  GCGGACTGGG  CAAAGCGGCA  CATTAAACAC  CGGTCGGCGG  CGAAAATGCT
351  GACGCTCGCT  CTGCTGTTGC  TAACCTTTAT  CGACGACIAT  TTCCACAGCC
401  TCGGCGCTCG  TGCAGATTGC  CGCCCGGTGA  CGACAGATT  TAAAGTTTCC
451  CGGCGCAAAC  TCGCCTACAT  CCTCGACTCC  ACTGCTCGCG  CCATGCGGCT
501  GCTGATGCC  GTTTCAGCT  GGGGCGCGTC  GATTATGCC  ACGCTTGCCG
551  GATTGCTCGT  TACCTACAAA  ATTACCGAAT  ACACGCGCAT  GGGGACGTTT
601  GTGCGCATGA  GCCTGATGAA  CTATTACGCG  CTGTTTGCCC  TGATTATGGT
651  ATTGCTCGTC  GCATGGTTCT  CCTTCGACAT  CGGCTCGATg  gCGCGTTTCG
701  AACAGGCTGC  GTTGAACGAA  gccacggacg  aaacgcgcgc  tTCAGACGCT
751  ACCAAGGCTC  GTGTTTACGC  ATTGATTATT  CCGCTTTTGG  CCTTAATGCG
801  CTCAACGGTT  TCCGCCATGA  TCTACACCGG  CGCGCAGGCA  AGCGAAACCT
851  TCAGCATTTT  GGGGCGATTT  GAAAATACCG  ACGTAACAC  TTCGCTGGTA
901  TTCGGCGGCA  CTTGGCGCGT  GCTTGCCGTC  GTCTCTGCA  CGTTCGGCAC
951  GATTAACACC  GCCGATTATC  CAAAGCCGT  GTGGCAGGGT  GCGAAATCCA
1001  TGTTTCGGCG  AATCGCCATT  TTAATCCTCG  CTGGCTCAT  CAGTACGTTT
1051  GTGCGGCAAA  TGCACACGGG  CGACTACCTC  TCCACGCTGG  TTGCGGGAAT
1101  CATCATCATC  GGTTCCTCTC  CCGTCATCTC  CTTCCTGCTC  CGACGAGTGA
1151  TGGCGTTTGC  CACAGGCACA  AGCTGGGGGA  CPTTCGGCAT  TATGCTGCGA
1201  ATTGCCGCCG  CCATGCGCGT  CAAAGTCGAA  CCGCGCTGTA  TTATcccgGT
1251  TATGTCCGCA  GTAATGGCGG  GGGCGGTATG  GGGCGACCAC  TGTTCGCCCA
1301  TCTCCGACAC  GACCATCCTG  TCGTCCACCG  GCGCGCGCTG  CAACCCATCT
1351  GACCACGTGA  CCTCGCAACT  GCCTTATGCC  CTGACGGTTG  CGCGCCGCGC
1401  CGCATCGGGC  TACCTCGCAT  TGGGCTCGAC  AAAATCCGCG  CTGTTGGGCT
1451  TTGGCAACGAC  CGGTATTGTA  TTGGCGGTGC  TGATTTTCT  GTTGAAGAT
1501  AAAAAACGCG  CCGACGTTTG  A

```

This encodes a protein having amino acid sequence <SEQ ID 696>:

```

1  MQLIDYSHSF  FSVVPFFLAL  ALAVITRRVL  LSLGIGILVG  VAVLVGPNV
35  DGLTHLKDMV  VGLAWADGDW  SLGKPKILVF  LILLGIFTSL  LTYSGSNQAF
101  ADWAKRHHKN  RCGAKMLTAC  LVFVTFIDDY  FHS LAVGAI  A  REVTDKFKVS
151  RAKLAYILDS  TASPMCVLMP  VSSWGSIIIA  ILAGLLVITYK  ILEYTPMGFT
201  VAMSLMNYIA  LFAIMIVFVV  AWESFDIGSM  ARFEQARLNE  AQDETAASDA
251  TKGRIYALIL  FVALILASTV  SAMIVYGAQA  SETESILGAF  ENTQVNVSLV
301  FSGTCVCLAV  VLCTFETIKT  ADYFKAVWQG  AKSMFGAIAI  LILAWLLSFV
351  VGEHMHTGDYL  STLVAGNIHP  GELPVILFLL  ASVMAFATGT  SWGTFGIMLP
401  TAAAMAVKVE  PALLIPCMSA  VMAGAVCGDH  CSPISDTTIL  SSTGARNCHI
451  DHVTSQLFYA  LTVAAAAASG  YLAGLTKESA  LLGFGTTGIV  LAVLIFLLKD
501  KKRADV*

```

ORF26ng and ORF26-1 show 98.4% identity in 505 aa overlap:

		10	20	30	40	50	60
orf26-1.pep		MQLIDYSHSFFSVVPPFFLALALAVITRRVLLSLGIGILVGVAVLVGPNVDGLTHLKDMV					
orf26ng		MQLIDYSHSFFSVVPPFFLALALAVITRRVLLSLGIGILVGVAVLVGPNVDGLTHLKDMV					
		10	20	30	40	50	60
orf26-1.pep		VGLAWSGDGWSLGKPKILVFLILLGIFTSLITYSGSNQAFADWAKRHHKNRGA	70	80	90	100	110
orf26ng		VGLAWADGWSLGKPKILVFLILLGIFTSLITYSGSNQAFADWAKRHHKNRGA	70	80	90	100	110
		130	140	150	160	170	180
orf26-1.pep		LVFVTFIDDYFHS LAVGAIARPVTDKFKVSRTKLAYILDSTAAAPMCVLMVPSW					
orf26ng		LVFVTFIDDYFHS LAVGAIARPVTDKFKVSRAKLAYILDSTASPMCVLMVPSW					
		130	140	150	160	170	180
		190	200	210	220	230	240

5	orf26-1.pep	TLAGLLVTKITEYTPMGTFVAMSLMNYALFALIMVFVWAFSFDIGSMARFEQAALNE
	orf26ng	TLAGLLVTKITEYTPMGTFVAMSLMNYALFALIMVFVWAFSFDIGSMARFEQAALNE
10	orf26-1.pep	AHDETAVS DATKGRVYALII PVLALIASTVSAMIYTGQAQASSETFSILGAFENTDVNTSLV
	orf26ng	AQDETAASDATKGRVYALII PVLALIASTVSAMIYTGQAQASSETFSILGAFENTDVNTSLV
15	orf26-1.pep	EGGTGCVLAVVLCTLGTIKTADYPKAVWQGAQSMFGAIAILILAWLISTVVGEMHTG DY L
	orf26ng	EGGTGCVLAVVLCTFGTIKTADYPKAVWQGAQSMFGAIAILILAWLISTVVGEMHTG DY L
20	orf26-1.pep	STLVAGNIHPGFLEPVILFLLASVMAFATGTSWGTFGIMLPFAAAMAVKVEPALIIPCMSA
	orf26ng	STLVAGNIHPGFLEPVILFLLASVMAFATGTSWGTFGIMLPFAAAMAVKVEPALIIPCMSA
25	orf26-1.pep	VMAGAVCGDHCSPISTDTILSSSTGARCNHIDHVTSQLPYALTVA AAAASGYALGLTKSA
	orf26ng	VMAGAVCGDHCSPISTDTILSSSTGARCNHIDHVTSQLPYALTVA AAAASGYALGLTKSA
30	orf26-1.pep	LLGFGTTGIVLAVLIFLTKDKKRANAX
	orf26ng	LLGFGTTGIVLAVLIFLTKDKKRADVX

In addition, ORF26 ng shows significant homology to a hypothetical *H. influenzae* protein:

```

sp|P4263|YF86_HAEIN  HYPOTHETICAL PROTEIN HI1586 >gi|1074850|pir|C64037
hypothetical
protein HI1586 - Haemophilus influenzae (strain Rd KW20) >gi|1574427 (U32832) H.
influenzae predicted coding region HI1586 [Haemophilus influenzae] Length = 519
Score = 538 bits (1370), Expect = e-152
Identities = 280/507 (55%), Positives = 346/507 (68%), Gaps = 7/507 (1%)

Query: 1 MQLIDYSHSFFSVVPPFLALALAVITRRXXXXXXXXXXXXAFLVGNPDVGLTHLTKDMV 60
M+LID+S S +S+VP LA+ LA+ TRR +L+ V
Sbjct: 14 MELIDFSSSVSVIPALLAIIAIAIATRRVLVLSAGIIIGSLMDQWIGSAFNYLVKNV 73

Query: 61 VGLAWADGDSWLSGPKPILVILLIGFTSLLTYSGSNQAFADWAKRHKIKRCKAKMLTAC 120
V L +ADG+ + I++FLLLG+ T+LLT SGN+AF+WA+ IK R GAK+L A
Sbjct: 74 VSLVYADGEIN-SNMNIVLFLLLGLVLTALLTVSGSNRAFAEWAQSRIRKGRGAKLLAAS 132

Query: 121 LVFVTFTDDYFHLAVGAIARPVTDKFKVSRAKLAYILDSTASPMCVLMPVSWGASIIA 180
LVFVTFTDDYFHLAVGAIARPVTD+FKVSRAKLAYILDSTA+PMCV+MPVSWGGA II
Sbjct: 133 LVFVTFTDDYFHLAVGAIARPVTDKFKVSRAKLAYILDSTAAPMCMVMPVSWGAYIIT 192

Query: 181 TLAGLLVTKITEYTPMGTFVAMSLMNYALFALIMVFVWAFSFDIGSMARFEQAALNE 240
+ GLL TY ITEYTP+G FVAMS MN+YA+F++IMVF VA+ESFDI SM R E+ AL
Sbjct: 193 LIGGLLATYSITEYTPIGAFVAMSMNFYAIFSIIMVFVAYFESFDIASMVREHKLALKN 252

Query: 241 AQDETAASDATKGRVYALII PVLALIASTVSAMIYTGQAQ----SETFSILGAFENTDVN 296
+D+ TKG+V LI+P+L LI +TVS MIYTGA+A + FS+LG FEINT V
Sbjct: 253 TEDQLEEEETGKTQGVNRNLLPILVLIATVSMMIYTGAEALAADGKVSFVLGTFTENTVVG 312

Query: 297 TSLVFGTGCVL--AVVLCTFGTIKTADYPKAVWQGAQSMFGAIAILILAWLISTVVGEM 354
TSLV GG C ++ ++ + +Y ++ G KSM G + +VGM
Sbjct: 313 TSLVVGFCFSIIISTLLIILDRQVSVPEVVRSWIVGIKSMGAIILFAWTINKKIVGEM 372

Query: 355 HTGDYLTSLVAGNIHPGFLEPVILFLLASVMAFATGTSWGTFGIMLPFAAAMAVKVEPALI 414
TG YLS+LV+GNI FLPVILF+L + MAF+TGTSGWGTFGIMLPFAAAMA P L+

```

Sbjct: 373 QTGKYLSSLVSGNIPMQFLPVILFVLGAAMAFSTGTSGWTFGIMLPAAAMAANAPELL 432
 Query: 415 IPCMSAVMAGAVCGDHCSPISTTILSSTGARNHIDHVTSSQXXXXXXXXXXXXXXXXXXXX 474
 +FC+SAVMAGAVCGDHCSPISTTILSSTGA+CNHIDHVT+Q
 Sbjct: 433 LFCLSAVMAGAVCGDHCSPISTTILSSTGAKCNHIDHVTQLPYAATVATATSIGIVV 492
 Query: 475 XXXKSALLGFGTTGIVLAVLIFLLKDK 501
 S L G F T + L V + I F + K +
 Sbjct: 493 GFTYSLAGFAATAVSLIVIIFAVKKR 519

Based on this analysis, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 83

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 697>:

1 ..AAGCAATGGT ATGCCGACGN .AGTATCAAG ACGGAARTGG TTATGCTCAA
 51 CGATGAGCCT GCCAAATATC TGACTTGGGA TGAAGCGGC CGATTAATCT
 101 CGACCTCCAC CATCAACGCA ACGGGTGGT TTTGGAGTGG
 151 TATGAAGATG GTTCTAAAA GAGCGAAT. GTTTATCAGG ATGACAGATT
 201 GGTCAGGAAA ACCCAGTGGG ATAAGATGG TTATTATCA GAAACCTGA

This corresponds to the amino acid sequence <SEQ ID 698; ORF27>:

1 ..KQWYADXSIK TEMVMVNDEP AKILTWDSEG RLLSELSIRH HQRNGVVLEW
 51 YEDGSKKSEX VYQDDKLVRK TQWDKGYLI EP*

Further work revealed the complete nucleotide sequence <SEQ ID 699>:

1 ATGAAAAAAT TATCTCGGAT TGTATTTC AACTGCTGTG TGGGTTTTTC
 51 GGCCGCTTGT CCGCGCAGA CTTATCTGTT TTATTTTAAAT CAGAACGGAA
 101 AGCTGACGGC GACGATGTCT TCTGCCGCTT ATATCAGGCA ATATAGTGTG
 151 GTGGCGGGTA TTGCGCACGC CAGGATTTT TATTATCCGT CGATGAAGAA
 201 ATATTCTGAA CCTTATATCG TTGCTTCAAC GCAAATCAAA TCTTTTGTGC
 251 CTACCCCTGCA AAGCGGTATG TTGATTTTGT GGCATTTTAA TGGTCAGAAA
 301 AAAATGGCGG GGGGCTTCAG CAAGGGTAAG CCGGACGGGG AGTGGGTCAA
 351 CTGGTATCCG AACGGTAAAA AATCTGCCGT TATGCCCTAT AAAAATGGCT
 401 TGAATGAGGG TACGGGATAC GCTATTACG GTACAGGGGG CAAGGAAGC
 451 GAATCTCCGT TATAGCAAGA TAAGCAACAG CGCGATGGA AGCAATGGTA
 501 TGCCGACGGC AGTATCAAGA CGAAATGGT TATGGTCAAC GATGAGCCTG
 551 CAAAATTTCT GACTTGGGAT GAAAGCGGCC GATTACTCTC GGAAGTGTCT
 601 ATCCGCCACC ATCAACGCAA CGGGGTGGTT TTGGAGTGGT ATGAAGATGG
 651 TTCTAAAAAG AGCGAAGCTG TTTATCAGGA TGACAAGTTG GTCAGGAAAA
 701 CCCAGTGGGA TAAGGATGGT TATTATTCG AACCCGTA

This corresponds to the amino acid sequence <SEQ ID 700; ORF27-1>:

1 MKKLSRIVFS TVLLGFSAAL PAQTVSVYFN QNGKLTATMS SAAVIRQYSV
 51 VAGIAHAQDF YFYSMKKYSE PYIVASTQIK SFVPTLQNG LILWHFNQK
 101 KMAAGGFSKGL PDGEWNNWYP NGKKSVMVPY KNLSSEGTGY RYRNGFKES
 151 EIQFKQNKAN GVMKQWYADG SIKTEMVMVN DEPAKILTWD ESGRLLSELS
 201 IRRHQNRNGVV LEWYEDGSKK SEAVYQDDKL VRKTQWDKDG YLIEP*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF27 shows 91.5% identity over a 82aa overlap with an ORF (ORF27a) from strain A of *N.meningitidis*:

orf27.pep 10 20 30
 KQWYADXSIKTEMVMVNDEPAKILTWDSEG

-421-

```

      ||||| :|||||
orf27a  LSEGTGKRYRNRGGKESEIQFKQNKANGVWKQWYADGNIKTEMVMVNDPAKILTWDESG
      140      150      160      170      180      190

5      orf27.pep  RLLSELSIRHHQRNGVVLWEYEDGSKKSEXYQDDKLVKRTQWDKDYGLIEFX
      |||||:|||||
orf27a  RLLSELSIRHHQRNGVVLWEYEDGSKKSEAVYQDDKLVKRTQWDKDYGLIEFX
      200      210      220      230      240

```

10 The complete length ORF27a nucleotide sequence <SEQ ID 701> is:

```

1  ATGAAAAAAT  TATCTCGGAT  TGTATTTTCA  ACTGTCCTGT  TGGGTTTTTC
51  GGC CGCTTTG  CCGGCGCAGA  NCTATTCTGT  TTATTTTAA  CAGAACGGGA
101 AACTGACGSC  GACGNTGTCT  TCTGCCGCNT  ATATCAGGCA  ATATAGTGTG
151 GCGGAGGGTA  TTGCGCAOCG  GCAGGANTTT  TANTATCOGT  CGATGAAGAA
201 ATATTCCGAA  CTTTATATCG  TTGCTTCAC  GCAATCAAA  TCTTTTGTGC
251 CTACCCGTGA  AAACGSGTATG  TTGATTTTGT  GGCATTTAA  NGGTGAGAA
301 AAAATGGCNG  GGGGCTTCAC  CAGGCTAAG  CCGAGCGGG  AGTGGTCAA
351 CTGGTATCCG  AAGGTAATAA  AATCTGCGGT  TATGCTCTAT  AAAATGGTT
401 TGAATGTAAG  TACGGGSGTNN  CGCTATTACC  GTAACGCGCG  CAGGNAAGC
451 GAAATCCAGT  TTAACAAGAA  TAAGGCARAC  GGCATGGA  AGCAATGGTA
501 TGCCGACGCG  AATATCAAAA  CGGAATAGGT  TATGGTCAAT  GATGAGCCTG
551 CCAAAATTCT  GACATGGGAT  GAAAGCGGTG  GATTACTCTC  GGAACGTGCT
601 ATCCATCATC  ATNAACGTAA  TGGAGTAGTC  TTAGATGGT  ATGAAGATGG
651 TTCTAAAAAG  ANTGAAGCTG  TTTATCAGGA  TGATAAGTTG  GTCAGGAAAA
701 CCCAGTGGGA  TAANGATGGT  TATTTAATCG  AACCTGA

```

This encodes a protein having amino acid sequence <SEQ ID 702>:

```

1  MKKLSRIVFS  TVLLGFSAL  PAQXYSVYFN  QNGKLTATXS  SAAYIRQYSV
51  AEGIAHAQXF  XYPSMKKYSE  PYIVASTQIK  SFVPTLQNGM  LILWHFXGQK
101 KMAGGFSKKG  PDGEWVNWYP  NGKKSAMVPI  KNLSEGTGX  RYRNRGGKES
151 EIQFKQNKAN  GVWKQWYADG  NIKTEMVMVN  DEPAKILTW  ESRLLSELS
201 IHHXNRNGV  LEWYEDGSKK  KEAVYQDDKL  VRKTQWDKDG  YLIEP*

```

ORF27a and ORF27-1 show 94.7% identity in 245 aa overlap:

```

      10      20      30      40      50      60
orf27a.pep  MKKLSRIVFSTVLLGFSALPAQXYSVYFNQNGKLTATXSAAAYIRQYSV AEGIAHAQXF
      |||||:|||||
orf27-1      MKKLSRIVFSTVLLGFSALPAQTYSVYFNQNGKLTATMSAAYIRQYSV VAGIAHAQF
      10      20      30      40      50      60

      70      80      90      100     110     120
orf27a.pep  XYPSMKKYSEPYIVASTQIKSFVPTLQNGMLILWHFXGQKKMAGGFSKGGKPDGEWVNWYP
      |||||:|||||
orf27-1      YYPMSKKYSEPYIVASTQIKSFVPTLQNGMLILWHFNGQKKMAGGFSKGGKPDGEWVNWYP
      70      80      90      100     110     120

      130     140     150     160     170     180
orf27a.pep  NGKKSAMVPIKNGLSEGTGXRYRNRGGKESEIQFKQNKANGVWKQWYADGNIKTEMVMVN
      |||||:|||||
orf27-1      NGKKSAMVPIKNGLSEGTGYRYYRNRGGKESEIQFKQNKANGVWKQWYADGSIKTEMVMVN
      130     140     150     160     170     180

      190     200     210     220     230     240
orf27a.pep  DEPAKILTWDESGRLLSELSIRHHQRNGVVLWEYEDGSKKSEAVYQDDKLVKRTQWDKDXG
      |||||:|||||
orf27-1      DEPAKILTWDESGRLLSELSIRHHQRNGVVLWEYEDGSKKSEAVYQDDKLVKRTQWDKDG
      190     200     210     220     230     240

orf27a.pep  YLIEFX
      |||||
orf27-1      YLIEFX

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF27 shows 96.3% identity over 82 aa overlap with a predicted ORF (ORF27ng) from

N.gonorrhoeae:

```

5      orf27.pep                               KQWYADXSIKTEMVMVNDEPAKILTWDESG      30
      orf27ng      LSEGTGYRYYRNGGKESEIQFKQNKANGVWKQWYADGSIKTEMVMVNDEPAKILTWDESG      193

10     orf27.pep      RLLSELSIRHHQRNGVVLWEYEDGSKKSEXYQDDKLVRKTQWDKDGYLEIEP      82
      orf27ng      RLLSELSIRHHKRNGVVLWEYEDGSKKSEAVYQDDKLVRKTQWDKDGYLEIEP      245

```

The complete length ORF27ng nucleotide sequence <SEQ ID 703> is:

```

1      ATGAAGAAAT      TATCTCGGAT      TGGTATTTTCA      ATCCTACTGT      TGGGTTTTTC
51     GGC CGCTTTC      CCGCGCAGAA      CCTATCTGT      TTATTTTAA      CAGAACGGGA
101    AACTGACGCG      GACGATGTCT      TCTGCCGCTT      ATATCAGGCA      ATATAGTGTG
151    CGGCGGGTA      TCGCACACGC      CGAGGATTTT      TATTATCCGT      CGATGAAGAA
201    ATATTCCGAA      CTTATATCG      TTGCTTCAAC      GCAAAATCAA      TCTTTGTGTC
251    CTACCTCGCA      AAACGATATG      TTGATTTTGT      GGCATTTAA      TGGTCAGAAA
301    AAAATGGCGG      GGGGCTTCAG      CAAGGTAAG      CCGACGGGG      AATGGGTCAA
351    CTGGTATCOG      AACGGTAAAA      AATCTGCGGT      TATGCCTTAT      AAAAATGGCT
401    TGAGTGAGGG      TACGGGATAC      CGTTATTACC      GTAACGGCGG      CAAGGAAAGC
451    GAAATCCAGT      TTAAGCAAAA      TAAGCGCAAC      GGGGTATGGA      AGCAATGGTA
501    TGGCGATGGA      AGTATCAAGA      CGGAATGGT      TATGGTCAAC      GATGAGCTCG
551    CCAAAATCT      GACTTGGGAT      GAAAGCGGCC      GATTACTTTC      GGAAGTGTCT
601    ATCCGCGACC      ATAAACGCAA      CGGGGTGGT      TTGGAGTGGT      ATGAAATGGG
651    TTCTAAAAAG      AGCGAGGCTG      TTTATCAGGA      TGCAATGTG      GTCAGGAAAA
701    CCAATGGGA      TAAGGATGGT      TATTTAATCG      AACCTGGA

```

This encodes a protein having amino acid sequence <SEQ ID 704>:

```

1      MKKLSRIVFS      IVLLGFSAL      PAQTYSVYFN      QNGKLTATMS      SAAYIRQYSV
51     AAGLAHQDF      YFSMKYSE      PYIVASTQIK      SFVPTLQNM      LLLWHFNGQK
101    RMAGGFSKKG      PDGERVWYF      NGKKSAMVYF      KNGLSEGTGY      RYYRNGGKES
151     EIOFKQNKAN      GVWKQWYADG      SIKTEMVMVN      DEPAKILTW      ESGRLLESL
201    IRHHKRNGVV      LEWYEDGSKK      SEAVYQDDKL      VRKTQWDKDG      YLIEF*

```

ORF27ng and ORF27-1 show 98.8% identity in 245 aa overlap:

```

35     orf27-1.pep      10      20      30      40      50      60
      orf27ng      MKKLSRIVFSIVLLGFSALPAQTYSVYFNQNGKLTATMSSAAYIRQYSVVAAGIAHAQDF
      10      20      30      40      50      60

40     orf27-1.pep      70      80      90      100      110      120
      orf27ng      YYPSMKKYSEPYIVASTQIKSFVPTLQNGMLLWHFNGQKMGAGGFSKGPDEGWNWYP
      70      80      90      100      110      120

45     orf27-1.pep      130      140      150      160      170      180
      orf27ng      NGKKSAMVYPKNGLSEGTGYRYYRNGGKESEIQFKQNKANGVWKQWYADGSIKTEMVMVN
      130      140      150      160      170      180

50     orf27-1.pep      190      200      210      220      230      240
      orf27ng      DEPAKILTWDESGRLLSELSIRHHQRNGVVLWEYEDGSKKSEAVYQDDKLVRKTQWDKDG
      190      200      210      220      230      240

55     orf27-1.pep      YLIEPX
      100     YLIEPX

```


orf27ng

YLIEFX

Based on this analysis, including the putative leader sequence in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF27-1 (24.5kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 17A shows the results of affinity purification of the GST-fusion protein, and Figure 17B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for ELISA, which gave a positive result, confirming that ORF27-1 is a surface-exposed protein and a useful immunogen.

Example 84

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 705>:

```

1  ATGAAATTTA  CCAAGCACCC  CGTCTGGGCA  ATGGCGTTCC  GCCCATTTTA
51  TTGCGTGGCG  GCTCTGTACG  GCGCATGTCT  CGTATTGCTG  TGGGGTTTCG
101  GCTACACGGG  AAGCGACAG  CTGTCCGGTT  TCTATTGGCA  CGCGCATGAG
151  ATGATTTCGG  GTTATGCGGG  ACTGGTCGTC  ATCGCCTTCC  TGTGACCGCG
201  CGTCGCCACT  TGAACGGGG  AGCGGCCAC  GGGGGCGGCG  GATCTGGGTC
251  GGCTTGACTA  TCTTTGGGCT  GGCTGCGCGG  ATTCGCGCCT  TTATCCCGGG
301  TTGGGGTGG  TGGCAAGCG  GATACTCGG  TACGCTGTCT  TTCTGGTACG
351  GCGCGGTGTG  CATGGCTTTG  CCGTTATCC  GTTCGCAGAA  TCAACGCAAC
401  TATGTTGCG  TGTTCGCGCT  GTTCGTCCTG  GTCGGCAGCG  ATGCGGCGTT
451  CCACGTCCAG  TGCACAACG  GCAACCTAGG  CGGACTCTTG  AGCGGATTGC
501  AGTCGGGCTT  GTGTATG

```

This corresponds to the amino acid sequence <SEQ ID 706; ORF47>:

```

1  MKFTKHPVWA  MAFRPFYSLA  ALYGALSVLL  WFGYGTGTHX  LSGFYWHAHE
51  MIWGYAGLVV  IAPLLTAVAT  WTGQPPIRGG  VLVGLTI FWL  AARIAAFIPG
101  WGSASGILG  TLFFWYGAVC  MALFVIRSQN  QRNYVAVFAL  FVLGGTHAAF
151  HVQLHNGNLG  GLLSGLQSL  VM

```

Further work revealed the complete nucleotide sequence <SEQ ID 707>:

```

1  ATGAAATTTA  CCAAGCACCC  CGTCTGGGCA  ATGGCGTTCC  GCCCATTTTA
51  TTGCGTGGCG  GCTCTGTACG  GCGCATGTCT  CGTATTGCTG  TGGGGTTTCG
101  GCTACACGGG  AAGCGACAG  CTGTCCGGTT  TCTATTGGCA  CGCGCATGAG
151  ATGATTTCGG  GTTATGCGGG  ACTGGTCGTC  ATCGCCTTCC  TGTGACCGCG
201  CGTCGCCACT  TGAACGGGG  AGCGGCCAC  GGGGGCGGCG  GTTCTGGGTC
251  GCTTGACTAT  CTTTGGGCTG  GCTGCGGGA  TTGCGCGCTT  TATCCCGGTT
301  TGGGGTGGCT  CGGCAAGCG  CATACTCGGT  ACGCTGTCTT  TCTGGTACGG
351  GCGCGGTGTG  ATGGCTTTTC  CCGTTATCCG  TTGCGGAAAT  CAACGCAACT
401  ATGTTGCGCT  GTTCGCGCTG  TTGCTCTTGG  CGGCGACGCA  TGGCGCGGTC
451  CACGTCCAGC  TGCACAACGG  CAACCTAGGC  GGACTCTTGA  GCGGATTGCA
501  GTCGGGCTTG  GTGATGGTGT  CGGGTTTAT  CGGCTGATTT  GGTACGCGGA
551  TTATTTCGTT  TTTTACGTCC  AAACGCTTGA  ATGTGCGCGA  GATTCCCGAT
601  CCGAAATGGG  TGGCGCAGGC  TTGCTGTGG  CTGCCCATGC  TGACTGCCAT
651  GCGATGTCGG  CAGCGTGGCT  GTGCGCGCTT  TTGCTCTTGG  TTTGCTCTTG
701  CGGACGGTGT  GATTCTTAC  GTGCGAGTGT  ACCGCTGCGT  GATTAACGCG
751  GTGTTCAAG  AGCCGATGCT  TTGCGGCGCT  ATCTGCTTTAC  ATCTGCTTTAC
801  CGGATTGCGG  CTGATTGCGG  TCGCGCGCTC  TTATTTCAAA  CCGCCTTTC
851  TCAATCTGGG  TGTGCATCTG  ATGCGGTCG  GCGGTATCGG  CGTCTGACTG
901  TTGGGATGTA  TGGCGGCTAC  CGCGCTTGT  CATACGGGCA  ATCCGATTTA

```

951 TCCGCCGCC AAGCCGCTT CCGTTGCGTT TTGGCTGATG ATGCGCGCAA
 1001 CCGCCCTCG TATGTTGCG GTATTTTCTT CCGGCATGCG CTACACGCAC
 1051 AGCATCGCA COTCTTCGGT TTGTTTTCGA CTGCGGCTTT TGGTGTATGC
 1101 GTGGAAGTAT ATTCTTGGC TGATTGCTCC GCGTTGCGAC GGCAGGCCCG
 1151 GTTGA

This corresponds to the amino acid sequence <SEQ ID 708; ORF47-1>:

1 MKFTKHPVMA MAFRPFYSLA ALYGALSVLL WGFYGTGTHE LSGFYWHAHE
 51 MINGYAGLVV IAFLLTAVAT WTGQPPTRGG VLVGLTIFWL AARIAAFIPG
 101 WGSASGILG TLFYWGAVC MALPVIQRN QRNYVAVFAL FVLGGTHAAF
 151 HVQLHNGNLG GLLSGLSGL VMVSGFIGLI GTRIIISFPTS KRLNVQIPIS
 201 PKWVAQASLW LPLMTAMLMA HGVLAWL SAV FAFAGVIFT VQVVRWYKYP
 251 VLKEMPLWIL FAGYLFITGLG LIAVGASYFK PAFNLGVHL IGVGIGLVLT
 301 LGMMAATALG HTGNPIYFPP KAVPVAFWLM MAATAVRMVA VFSSGTAYTH
 351 SIRTSSVLFA LALLVYAWKY IPWLIRPRSD GRPG*

Computer analysis of this amino acid sequence predicts a leader peptide and also gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF47 shows 99.4% identity over a 172aa overlap with an ORF (ORF47a) from strain A of *N.*

meningitidis:

orf47.pep	10	20	30	40	50	60
orf47a	MKFTKHPVWMAFRPFYSLAALYGALSVLLWGFYGTGTHXLSGFYWHAHEMIWGYAGLVV					
	10	20	30	40	50	60
orf47.pep	70	80	90	100	110	120
orf47a	IAFLLTAVATWTGQPPTRGGVLVGLTIFWLAARIAAFIPGWGASASGIIGTLFFWYGAVC					
	70	80	90	100	110	120
orf47.pep	130	140	150	160	170	
orf47a	MALPVIQRNQRNYVAVFALEVLGGTHAAHFVQLHNGNLGGLSLGSLVM					
	130	140	150	160	170	180
orf47a	MALPVIQRNQRNYVAVFALEVLGGTHAAHFVQLHNGNLGGLSLGSLVMVSGFIGLI					
	190	200	210	220	230	240
orf47a	GTRIIISFPTSKRLNVQIPSPKVAQASLWLEMLTAMIMAHGVMPWLSAAFAFAAGVIFT					

The complete length ORF47a nucleotide sequence <SEQ ID 709> is:

1 ATGAATTTA CCAAGCACC CGTTTGGGCA ATGCGCTTCC GCCCGTTTTA
 51 TTCACTGCGG GCTCTGTACG GCGCATTTGC CGTATTGCTG TGGGGTTTCG
 101 GCTACACGGG AACGCACGAG CTGTCCGGTT TCTATTGGCA CGCGCATGAG
 151 ATGATTGGG GTTATGCCGG ACTGGCTGTC CTGCGCTTCC TGCTGACCGC
 201 CGTCGCGCACT TGACGCGGCG AGCGCGCCAC GCGGGCGCGC GTTCTGGTGC
 251 GCTTGACTAT CTTTGGCTG GTCGCGCGCT TATCGCGCTT TATCGCGCTT
 301 TGGGCTGCGT CGGCAAGCGG CATACCTCGG ACTCGCTGTT TCTGGTACGG
 351 CGCGGTGTGC ATGCGTTTGC CCGTTATCCG TTCGAGAT CAACGCAATT
 401 ATGTTGCGCT GTTCCGCGCTG TTGCTCTGG CGGATACGCA CGCGCGGCTC
 451 CACGTCACCG TGCACACGCG CAACCTAGGC GGACTCTTGA GCGGATTGCA
 501 GTCGGGCTTG GTGATGGTGT CGGGTTTTAT CGGTCTGATT GGTACGCGGA
 551 TTATTTCGTT TTTTACGTCC AAACGGTTGA ATGCGCGCA GATTCCCGAT
 601 CCGAATGGG TGGCGCAGCG TTGCTGTGG CTGCGCATCG TGACCGCCAT
 651 GCTGATGGCG CACGCGCTGA TGCCTTGGCT GTGCGCGCTT TTGCGGTTTG
 701 CGGCAAGTGT GATTTTTACC GTGCAGGTGT ACCGCTGTGT GTATAAGCTT
 751 GTGTTGAAAG AGCCGATGCT GTGGATCTCG TTGCGCGGCT ATCTGTTTAC
 801 CGGATTGGGG CTGATTGCGG TCGCGCGCTC TATTTTCAAA CCGGCTTTCC
 851 TCAATCTGGG TGTGCATCTG ATCGGGGTCG GCGGTATCGG CGTGCTGACT

5 901 TTGGGCATGA TGGCGCGTAC CGCGCTCGGT CATAACGGCA ATCCGATTGA
 951 TCCGCCGCC AAGCCGCTC CCGTGCCTT TTGGCTGATG ATGCGGGCAA
 1001 CCGCCGTCG TATGTTTGC GTATTTTCTT CCGGCACTGC CTACACGCAC
 1051 AGCATACGCA CCTCTTCGGT TTTGTTTGA CTCGCGCTT TTGTTGATGC
 1101 GTGGAAGTAT ATTCCTTGGC TGATTCGTC GCGTTCGGAC GGCAGGCCCG
 1151 GTTGA

This encodes a protein having amino acid sequence <SEQ ID 710>:

1 MKFTKHPVWA MAFRPFYSLA ALYGALSVLL WGRGYTGTHE LSGFYWHAHE
 51 MIWGYAGLVV IAFLLTAVAT WTGQPPTRGG VLVLTIIFWL AARIAAFIPG
 101 WGAASAGILG TLFFWYGAVC MALPVRISQN QRNYVAVFAL FVLGGTHAAE
 151 HVQLHNGNLG GLSLGQSLG VMVSGFIGLI GTRIIISFFTS KRLNVQPIPS
 201 PKWVAQASLW LPMLTAMLMA HGVMPWLSAA FAFVAGVIFT VQVYRWYKYP
 251 VLKEPMLWIL FAGYLFGLG LIAGVASYFK PAFNLGVHIL IGVGIGIVLT
 301 LGMMARTALG HTGNPIYPPP KAVPFAFWLM MAATAVRMVA VFSSGTAYTH
 351 SIRTSSVLFA LALLYAWKY IPWLIRPRSD GRG*

ORF47a and ORF47-1 show 99.2% identity in 384 aa overlap:

		10	20	30	40	50	60
orf47a.pep		MKFTKHPVWMAFRPFYSLAALYGALSVLLWGRGYTGTHELSGFYWHAHEMIWGYAGLVV					
orf47-1		MKFTKHPVWMAFRPFYSLAALYGALSVLLWGRGYTGTHELSGFYWHAHEMIWGYAGLVV					
		10	20	30	40	50	60
orf47a.pep		IAFLLTAVATWTGQPPTRGGVLVGLTIIFWLAARIAAFIPGWGASAGILGTLFFWYGAVC					
orf47-1		IAFLLTAVATWTGQPPTRGGVLVGLTIIFWLAARIAAFIPGWGASAGILGTLFFWYGAVC					
		70	80	90	100	110	120
orf47a.pep		IAFLLTAVATWTGQPPTRGGVLVGLTIIFWLAARIAAFIPGWGASAGILGTLFFWYGAVC					
orf47-1		IAFLLTAVATWTGQPPTRGGVLVGLTIIFWLAARIAAFIPGWGASAGILGTLFFWYGAVC					
		70	80	90	100	110	120
orf47a.pep		MALPVRISQNQRNYVAVFALFVLGGTHAAHFVQLHNGNLGSLGSLGVLVMSGFIGLI					
orf47-1		MALPVRISQNQRNYVAVFALFVLGGTHAAHFVQLHNGNLGSLGSLGVLVMSGFIGLI					
		130	140	150	160	170	180
orf47a.pep		MALPVRISQNQRNYVAVFALFVLGGTHAAHFVQLHNGNLGSLGSLGVLVMSGFIGLI					
orf47-1		MALPVRISQNQRNYVAVFALFVLGGTHAAHFVQLHNGNLGSLGSLGVLVMSGFIGLI					
		130	140	150	160	170	180
orf47a.pep		GTRIIISFFTSKRLNVQPIPSPKWVAQASLWLPMLTAMLMAHGVMPWLSAAFAVAGVIFT					
orf47-1		GTRIIISFFTSKRLNVQPIPSPKWVAQASLWLPMLTAMLMAHGVMPWLSAAFAVAGVIFT					
		190	200	210	220	230	240
orf47a.pep		GTRIIISFFTSKRLNVQPIPSPKWVAQASLWLPMLTAMLMAHGVMPWLSAAFAVAGVIFT					
orf47-1		GTRIIISFFTSKRLNVQPIPSPKWVAQASLWLPMLTAMLMAHGVMPWLSAAFAVAGVIFT					
		190	200	210	220	230	240
orf47a.pep		VQVYRWYKYPVLKEPMLWILFAGYLFGLGLIAGVASYFKPAFNLGVHILIGVGIGIVLT					
orf47-1		VQVYRWYKYPVLKEPMLWILFAGYLFGLGLIAGVASYFKPAFNLGVHILIGVGIGIVLT					
		250	260	270	280	290	300
orf47a.pep		VQVYRWYKYPVLKEPMLWILFAGYLFGLGLIAGVASYFKPAFNLGVHILIGVGIGIVLT					
orf47-1		VQVYRWYKYPVLKEPMLWILFAGYLFGLGLIAGVASYFKPAFNLGVHILIGVGIGIVLT					
		250	260	270	280	290	300
orf47a.pep		LGMMAARTALGHTGNPIYPPPKAVPFAFWLMAATAVRMVAVFSSGTAYTHSIRTSSVLFA					
orf47-1		LGMMAARTALGHTGNPIYPPPKAVPFAFWLMAATAVRMVAVFSSGTAYTHSIRTSSVLFA					
		310	320	330	340	350	360
orf47a.pep		LGMMAARTALGHTGNPIYPPPKAVPFAFWLMAATAVRMVAVFSSGTAYTHSIRTSSVLFA					
orf47-1		LGMMAARTALGHTGNPIYPPPKAVPFAFWLMAATAVRMVAVFSSGTAYTHSIRTSSVLFA					
		310	320	330	340	350	360
orf47a.pep		LALLYAWKYIPWLIRPRSDGRPGX					
orf47-1		LALLYAWKYIPWLIRPRSDGRPGX					
		370	380				
orf47a.pep		LALLYAWKYIPWLIRPRSDGRPGX					
orf47-1		LALLYAWKYIPWLIRPRSDGRPGX					
		370	380				

Homology with a predicted ORF from *N.gonorrhoeae*

60 ORF47 shows 97.1% identity over 172 aa overlap with a predicted ORF (ORF47ng) from *N.gonorrhoeae*:

ORF47 MKFTKHPVWMAFRPFYSLAALYGALSVLLWGRGYTGTHELSGFYWHAHEMIWGYAGLVV 60

	ORF47ng	MKTTKHFWVMAAFRFFPYSLAALYGALSVLWLWGFSTGTTHLSGPYWHAHIMINGVAGLVV	60
5	ORF47	IAPLLTAVATWTGQPPTRGGVLGLITFLWLAARIAAIPGWGASASGLLGLTFEYWGAVC	120
	ORF47ng	IAPLLTAVATWTGQPPTRGGVLGLTAWFLAARIAAIPGWGAASGLLGLTFEYWGAVC	120
	ORF47	MALPVIQRNQNRNYAVFAFLVLGGTHAAHFVQLHNGNLGGLLSGLQSCLVM	172
10	ORF47ng	MALPVIQRNQNRNYAVFAIFVLGGTHAAHFVQLHNGNLGGLLSGLQSCLVMVGFTGLI	180

The ORF47ng nucleotide sequence <SEQ ID 711> is predicted to encode a protein comprising amino acid sequence <SEQ ID 712>:

1 MKF¹TKHPVWA MA²RPFPYS³LA ALYGALS⁴VLLW WGFYGTOT⁵HE LSGF⁶FWHAHE
51 MIWAGYAGLV⁷W IAF⁸LFL⁹TVAT¹⁰ WTGQ¹¹PTTRGG LRVL¹²GTAFWL AARIAA¹³FIPG
15 101 GAAAGAGSL¹⁴G L¹⁵TFPFWYAG¹⁶ ML¹⁷PVTRSN RNNV¹⁸VAFAL FVLG¹⁹THAFA
151 HVQLHNGNL²⁰G GLLSGLQSL²¹ VMVWGFI²²GL GMKIIS²³FFTS KRLKLP²⁴QIPS
201 PKWVHASLW LPLNAILMA²⁵ HRVMP²⁶LSAA PFFAAG²⁷VIFT VQVYAG²⁸GITP
251 IETFS²⁹CGSVA GIC³⁰RYLGNSS G

The predicted leader peptide and transmembrane domains are identical (except for an Ile/Ala substitution at residue 87 and an Leu/Ile substitution at position 140) to sequences in the meningococcal protein (see also *Pseudomonas stutzeri* orf396, accession number e246540):

TM segments in ORF47ng				
25	INTEGRAL	Likelihood = -5.63	Transmembrane	52 - 68
	INTEGRAL	Likelihood = -3.88	Transmembrane	169 - 185
	INTEGRAL	Likelihood = -3.08	Transmembrane	82 - 98
	INTEGRAL	Likelihood = -1.91	Transmembrane	134 - 150
	INTEGRAL	Likelihood = -1.44	Transmembrane	107 - 123
	INTEGRAL	Likelihood = -1.38	Transmembrane	227 - 243

Further work revealed the complete gonococcal DNA sequence <SEQ ID 713>:

30	1	ATGAAATATG	GCACAACTATG	CGCATCTGGGCA	ATGGCGCTGTG	GCCCGTTTTTA
	51	TTCACATGGCG	GCACATGTACG	GGCAATTGTG	CGTATGTCTG	TGGGGTTTTC
	101	GCTACACGGG	AACGACACGG	CTGCTCGGGT	TCTATTGGCA	CGCGCATGAG
	151	ATGATTTCGG	GTTATCCGAG	TCTGTCTGTC	TACGCGTTTC	CTGTCACCG
	201	CGTCCGCAT	TGGACGGGAC	AGCGCCCGAC	GAGGGGGCGG	GTTCTCGGGT
35	251	GCTTGCACGC	CTTTTGGCTG	CGTCGGCGAC	TTCGCCGCTT	TATCCCGGGT
	301	TGGGGTGGCG	CGGCAAGCGG	ACTACATCGT	ACGCTGTGTT	TCTGTCAGTC
	351	CGGGGTGTGC	ATGGCTTTGC	CGGTTATCCG	TctGCAAAAC	CGCGCAACT
	401	ATGctgcCGT	ATTCCGAATC	TTTGTGCTGG	CGGGTACGTC	TGGGggtTTC
	451	CGCctccAgc	TGCACACGG	CAACCTGAGC	GAGCTCTTGA	CGCGATGTCA
40	501	GTGCGGCTGC	CGTATGCTGT	CGGCTCTTTC	CGGCTCTTTC	CGGCTCTTTC
	551	TTATTTCGTT	TTTACGCTTC	AAACGGTTGA	ACGTTCCGCA	GATTCCCACT
	601	CGGAATGGT	TGGCGCAGCG	TTCGCTGGTC	TACACCATC	TGACCGCCAT
	651	ACTGATGGCG	CACGGGTGTA	TGCTCTGGCT	TCGGCGGGT	TTCCGGCTTTG
	701	CGGCGGGCGT	GATTTTATCC	GTACAGGTGT	ACCGCTGGTG	GATATAAACCT
45	751	GTATTGAAAG	AACCGATGCT	TGGGATTCTG	TTTTCGGCGT	ATCTGTTTAC
	801	CGGATTTGGG	CTGATTTGGG	TGCGGCGCTC	TTATTTCAA	CCTGCTCTTC
	851	TCAATCTGGG	CGTACATCTG	ATTCCGGGCT	CGGGTATTCG	CGTCTGACT
	901	TGGGCATGA	TGGCGCTATC	CGGCGTCGAT	TACAAGGCA	ATTCTGATTT
	951	TCCGCGCGCC	ACAGCCGCTC	CGGTTTCGTT	TGGCGGTATG	GTACGCGCAA
50	1001	CGCGTTCG	TATGCTTTCG	CGGCTCTTTC	CGGCTCTTTC	CGGCTCTTTC
	1051	AGCATACGCG	CTTCTCTGCT	CTATTTTCTA	CGGCTCTTTC	CGGCTCTTTC
	1101	GTGGAAATAC	ATTCCGTGGC	TGATCGCTCC	CGGTTCCGAC	GGCAGGCCCG
	1151	GTTGA				

This encodes a protein having amino acid sequence <SEO ID 714: ORF47ng-1>:

55 1 MKFTKHPVWA MAFRPFYSLA ALYGALSULL WGFYGTGTHE LSGFYWHAHE
51 MWGYAGLVV IAFLLTAVAT WTGQPPTRGG VLVGLTAEWL AARIAAFIPG
101 WGAAASGILG TLFFWYGAVC MALPVIRSON RRNYVAVFAI EVLGSTHAA

151 HVQLHNGNLG GLLSGLQSLG VMVSGFIGLI GMRIISFFTS KRLNVQPISP
 201 PKWVAQASLW LPLMTAILMA HGVMWPLWSA FAFAGVIFT VQVYRWYKYP
 251 VLKEPMLWIL FAGYLFTEGL LIAVGASYFK PAFNLGVHL LGVGGIGVLT
 301 LGMMARTALG HTGNSIYPPK KAVPVAFWLM MAATAVRMVA VFSSGTAYTH
 351 SIRTSSVLFA LALLVYAWKY IPWLIRPRSD GRGP*

ORF47ng-1 and ORF47-1 show 97.4% identity in 384 aa overlap:

		10	20	30	40	50	60
orf47-1.pep		MKFTKHPVMAFRPFYSLAALYGALSVLLWGFYTGTHELSGFYWHAHEMIWGYAGLVV					
orf47ng-1		MKFTKHPVMAFRPFYSLAALYGALSVLLWGFYTGTHELSGFYWHAHEMIWGYAGLVV					
		10	20	30	40	50	60
		70	80	90	100	110	120
orf47-1.pep		IAFLLTAVATWTGQPPTRGGLVGLTIFWLAARIAAFIPGWGASASGILGTLFFWYGAVC					
orf47ng-1		IAFLLTAVATWTGQPPTRGGLVGLTIFWLAARIAAFIPGWGASASGILGTLFFWYGAVC					
		70	80	90	100	110	120
		130	140	150	160	170	180
orf47-1.pep		MALPVIRSONRNYVAVFAFVLGGTHAAFHVOLHNGNLGGLSLGSLVMVSGFIGLI					
orf47ng-1		MALPVIRSONRNYVAVFAFVLGGTHAAFHVOLHNGNLGGLSLGSLVMVSGFIGLI					
		130	140	150	160	170	180
		190	200	210	220	230	240
orf47-1.pep		GTRISFFTSKRLNVQPISPFWKVAQASLWLPMLTAMLAHGVLAWSAVFAFAAGVIFT					
orf47ng-1		GTRISFFTSKRLNVQPISPFWKVAQASLWLPMLTAMLAHGVLAWSAVFAFAAGVIFT					
		190	200	210	220	230	240
		250	260	270	280	290	300
orf47-1.pep		VQVYRWYKYPVLKEPMLWILFAGYLFTEGLLIAVGASYFKPAFNLGVHLIGVGGIGVLT					
orf47ng-1		VQVYRWYKYPVLKEPMLWILFAGYLFTEGLLIAVGASYFKPAFNLGVHLIGVGGIGVLT					
		250	260	270	280	290	300
		310	320	330	340	350	360
orf47-1.pep		LGMMARTALGHTGNSIYPPPKAVPVAFWLMMAATAVRMVAVVFSSGTAYTHSIRTSSVLFA					
orf47ng-1		LGMMARTALGHTGNSIYPPPKAVPVAFWLMMAATAVRMVAVVFSSGTAYTHSIRTSSVLFA					
		310	320	330	340	350	360
		370	380				
orf47-1.pep		LALLVYAWKYIPWLIRPRSDGRGPX					
orf47ng-1		LALLVYAWKYIPWLIRPRSDGRGPX					
		370	380				

Furthermore, ORF47ng-1 shows significant homology to an ORF from *Pseudomonas stutzeri*:

gn1|PIDe246540 (273914) ORF396 protein [Pseudomonas stutzeri] Length = 396
 Score = 155 bits (389), Expect = 5e-37
 Identities = 121/391 (30%), Positives = 169/391 (42%), Gaps = 21/391 (5%)

Query: 7 PFWMAFRPFYSLAALYGALSVLLWGFYTGTHELSGFY-----WHAHEMIWGYAGLV 59
 P+W +AFRPF+ +LY L++ LW +TG GF WH HEM++G+A +

Sbjct: 14 PIWRLAFRPFPLAGSLYALLAIPLWVAWNTGLWF--GFQPTGCGWLAWHRHMLFGFAMAI 71

Query: 60 VIAFLLTAVATWTGQPPTRGGLVGLTIFWLAARIAAFIPGWGASASGILGTLFFWYGAV 119
 V FLLTAV TWGQ G LVGL A WLAAR+ ++G AA L LF

Sbjct: 72 VAGFLLTAVQTWGTQTAPSGNRLVGLAAVWLAARL--GWLFGFLPAWLAFLDLLFLVALVW 130

Query: 120 CMALPVIRSONRNYVAVFAFVLGGTHAAFXXXXXXXXXXXXXXXXXXXXXXVSGFIGLI 179
 MA + + +RNY V + + + G + + + + L

Sbjct: 131 MMAQMLWAVRQKRNYPVVVLSMLGADVLIITLGLQNDALQRQSVLAGLWLAALMAL 190

Query: 180 IGMRIISFFTSKRLNVQPISP-KWVAQASLWLPMLTAILMAHG-----MPWLSAFAFA 234
 IG R+I FFT + L P W+ A L + + + L A GV P L F A

Sbjct: 191 IGGRVIFPFTQRLGKVDVAVKPNWLDVALLVGTGVIALLHAFGVAMRPQPLGLLGFV-A 249

Query: 235 AGVLEFVQVYRWYKPVLEKPMILFAGLYFTGLGLIAGVASYF-KPAFXKXXXXXXXXXX 293
GV +++ RW+ K + K +LW L L+ + + +F A

5 Sbjct: 250 IGVGHLRLMRWYDKGIWKVGLLWSLHVAMLVVVAAGFLALWHFGLLAQSSPSLHALSV 309

Query: 294 XXXXXXXXXXXMMARTALGHTGNSIYPPPKAVPVFAFWLXXXXXXXXXXXXFSSGTAYTHSIR 353
M+AR LGHTG + P + AF L F S +

10 Sbjct: 310 GSMGLILAMIARVTLGHTGRPLQLPAGIIG-AFVL---FNLGTAARVFLSVAWPVGGLW 365

Query: 354 TSSVLFALALLVYAWKYIPWLIRPRSDGRPG 384
++V + LA +Y W+Y P L+ R DG PG

Sbjct: 366 LAACWTLAFALVFWRYAPMLVAARVDGHPG 396

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 85

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 715>:

```

1  ..ATGCCGCTCTG AAGGTTCAAG CGGCmTCGGT GyCGGGGAAY CAGAAGyGST
51 AGCGCATGCC CAATGAGACT TCGTGGGTTT TGAAGCGGST GTTTTCCAAG
101 CGTCCCAAGT TGTGGTAACG GTATCCGGTG TCyAArGTCA GCTTGGGyGT
151 GATGTCGAAA CCGACACCGG CGATGACACC AAGACCyAmG CTGCTGATrC
201 TGTkGCTTTC GTGATAGGSa GCTTTGyTGG kmksAsyTTG TAyrATwkkG
251 CCTssCwsTG kAGmGCKtk CkyTGGTtkA swGwArTAG TCGTGGTtTy
301 TkTtTyCACc GAATGAACyT GATGTTTAAc GTGTCCGTAG GCGACGCGCG
351 CGCCGATATA GGGTTTGAAT TTAzCGTTGA GTTTGAATTC GTAAATGGCG
401 GACAAGCCGA GAGAAGAAAC GCGGTGGAAG CTGCGGTTTC CCGATGTTT
451 TGTTTGGGTT TCtTTGTAGT TGtGTtTAT CTCTTCAGTA ACTTtTTTAT
501 TAGAAGAATr ACtTTCTTTC CATtTtCTGT AACTtGCATA ATcTGCCGCT
551 ATTCTCCAGC CGCCGAATTC ..

```

This corresponds to the amino acid sequence <SEQ ID 716; ORF67>:

```

1  ..MPSEGSdGXG XGEKEXVAHA QXDFVGFAG VFQASPVVVT VSGVXXQLGX
51 DVETDTGDDT KTXAADXVAF VIGRFKXKXL YXXAXKXKXK XWXXKXSRGF
101 XXHRMNLMPN VSVGDRADDI GFEEFIVEFEI VNGGQAEERN GVEAAVSLMF
151 CLGFFVVVVVY LFSNFFSRRRI TFFPFVSVTGI ICRYSPAEAI ..

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.gonorrhoeae*

ORF67 shows 51.8% identity over 199 aa overlap with a predicted ORF (ORF67ng) from *N.gonorrhoeae*:

orf67.pep		MPSEGSdGXG XGEKEXVAHA QXDFVGFAG	30
orf67ng	TNFEIAVLSGMTVRVFCARPAPVNGGRKMPSEGSdGIGIGSEAVAHAAQRFVGFAG	146	
	90 100 110 120 130 140		
orf67.pep	VFQASPVVVTVSGVXXQLGXDVETDTGDDTKTXAADXVAFVIGRFKXKXL YXXAXKXKXK	90	
orf67ng	VFQASPVVVAAGVQCGQAGRDVYAHARHRAEAQAQAAVAFILIGVFLRMSVRINRNNCCVSI	206	
orf67.pep	XWXXKXSRGFXXHRMNLMPNVSVDGARADIGFEFIVEFEIVNGGQAEERRNGVEAAVSLMF	150	
orf67ng	TRVGKSTCYFFSRIDAVSDVSVGDARTDIGFEFVFEIVNGGQAEERRNGVECAVFLMF	266	

```

orf67.pep      CLGFFVV-----VVYLFSNFFSRITFF-PFSVTGIICRYSPA AEI      190
               | | |           :: | : | : | : | | |||| | :|||:
orf67na       RLLVFYVKLVAAKSFIILSEQLFYVHGIFIVVPPVPTGIIRGDAPAAEVVADRHGPGVDGM 326

```

The ORF67ng nucleotide sequence <SEQ ID 717> is predicted to encode a protein comprising

5 amino acid sequence <SEQ ID 718>:

10

1	MPSETVGSIV	NGVGVDSVGF	SPFPFSIQHF	YFRCHTHRI	LRPPQPMQOL
2	NRHSHSGSNL	GRGVWATVSI	DKFPCQGVRI	PACRAGTNFE	IAVLSGMTVRI
101	VYFCARPAFV	QCGRLKMPSE	GSDDIGISGF	EVAVAHACRGF	VFEAGFAGVRA
151	SPVVVAIVAV	QCQGRALVRI	HARHIREAQQA	AAVAVALIGV	FLRMSVIRNA
201	NCCVSTIVRG	AKSTCYFFSR	IDAVSVLVSQ	DARTDITGFF	VVEFEIVNGQ
251	QAERRNGVEG	VGFIMFRLLV	FYVKLVAASK	FILTSOLFVY	VHGIFIVVPF
301	PVTGITRGDA	PAEAVVADRH	PGVDGMRDVI	SEITAYRAYF	VFAWSGWFRI
351	IVGNAEIGVG	*			

Based on the presence of a several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 86

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 719>

20	1	ATGTTTGGCTT	TTTGTGAAGC	CTTTTTCGTG	GAATACGGTT	ATFGCGGCTGT
	51	TTTCTTTTGTG	TTGTGACATG	CGGTTCATCT	CGCGAGGATT	CCCGAGGATT
	101	TGACCTTGGT	ACACAGCGGC	GTGATTTCGG	GATGCGGTTA	TACCAATCCG
	151	CATATTATGT	TTTCAGT CGG	TATGCTCCGG	CATATTGGTG	GGGACGGCAT
	201	CATGTTGCCG	CCTCAGCAGAA	TTCGCGGGCA	GAATATCTTA	rgGTTCAEAC
25	251	CTATTCCGSG	CATCATCAGC	CGGTACGGTT	ATGACGAGAT	TACGAGAAAA
	301	TTGCGACGTC	CGCCACAGTC	GGCTCATAGT	TTGCGCGGAT	TTCGCGCGG
	351	TTTGTGAAGC	GATCATATTG	TATACGCGGG	TATACGCGGG	AAAGTTCCTAT
	401	ACTTGGCGTT	TCCTATTATG	GTATACGCTG	CGCGA...	

This corresponds to the amino acid sequence <SEQ ID 720; ORF78>:

30 1 MFAPLEAFFV EYGYAAVEFV LVICGFGVPI PEDLTLTGG VISGMYGNTN
51 HMFVAVGLG VLVGDGIMFA AGRIWGQXXL XFXPIAXIMT PXRYEQVQEK
101 FDKYGNWVLF VARFLPLGIRT AVFVTAGISR KVSYLRFIIM DGLAA...

Further work revealed the complete nucleotide sequence <SEQ ID 721>:

35	1	ATGTTTGGCTT	TTTTAGAGCG	CTTTTGTTCG	GAATACGGTT	ATCCGGCGCTG
	51	TTTTTTTGTA	TGTGTCATAT	CGGTGTCATG	CGTCGCGAAT	COTGAGGAGTT
	101	TGACCTTCGT	ACAGCGCGCG	GTGATTTCCG	GATATGGGTTA	TACCAATACAT
	151	CATATTATGT	TGTGCATGTC	TATGCTCGCG	GATTTGTGTG	GGGACGGCAT
	201	CATGTTCCGC	CCTGACGAA	TTTTGGGGCA	GAATACTCTA	AGGTTCAACAC
	251	CTATTCGCG	CATATGCGC	CGGAACAGCT	ATGACAGGT	TCAGGAAAAA
40	301	TTTCGACAAAT	AGCGTAACGT	GGTCTATTG	GTTCGCGGTT	TCTCGCCGCT
	351	TTTGAGAAGC	GCGGTATTG	TATCGACGCG	GATCGCGC	AGAGGTTTCAT
	401	ACTTTCGGTT	TATCATATG	ATAGACGTGC	CGCGACTATG	TTCGTCCTCT
	451	ATTTCGATT	TATCGGGCA	ATACGGTGG	CACACATCG	ATTGGCTGAT
	501	GGCGAAATG	CACAGCTCG	AATCGGATG	TTTTGTTATC	TTGGGTATAG
	551	GTCGACCGT	TGTCGCTGG	ATTTCGGTGA	AAAAACGCA	ACGTATCCAG
45	601	TTTTACCGCA	CGAAATGAA	AGAAAGCGG	GGCGACGCA	AAGCGCCCAA
	651	GGCAGCCBAA	AAAGCCGCG	AAAGCAACCA	ATGA	

This corresponds to the amino acid sequence <SEQ ID 722; ORF78-1>:

50

1	MFAFLEAFFV	EYGYAAVFFV	LVICGFGVPI	PEDLTlVTGG	VISMGYTNP
51	<u>HIMFAVGMLG</u>	<u>VLVGDGINFA</u>	AGRIWGQKIL	RFKPIARIMT	PKRYEQVQEK
101	FDKYGNWVLF	VARELPGLTR	AVFVTAGISR	KVSYLRFITM	DGLAALISVP
151	IWIYLGGEYA	HNDIWLMAKM	HSLSQSGIFVI	LGITAGVVAW	IWWKKRQRIQ

201 FYRSKILKEKR AQRKAAKAAK KAAQSKQ*

Computer analysis of this amino acid sequence predicts several transmembrane domains, and also gave the following results:

Homology with the dedA homologue of *H.influenzae* (accession number P45280)

- 5 ORF78 and the dedA homologue show 58% aa identity in 144aa overlap:

```

Orf78: 4   FLEAFFVEYGYAAVFFVLVICGFGVPIDPDLTLVTGGVISGM--GYTNPHIMFAVGMGLV 61
          FL FF EYGY AV FVL+ICGFGVPIDP+TLV+GGVI+G+      N H+M V M+GV
DedA:  20   FLIGFFTEYGYWAVLFLVICGFGVPIDITLVSGGVVAGLYPENNVSHMLLVSMIGV 79

10 Orf78: 62  LVGDGIMFAAGRIWGQXXLFXPIAXIMTPKRYEQVQEKFDKYGNNWVLFVARFLPGLRTA 121
          L GD M+ GRI+G L F PI I+T R V+BKF +YGN VLEVARFLPGLR
DedA:  80   LAGDSMCWYLGRIYGTKILFRPIRRIVTLQRLMVRKEFSQYGNRVLFVARFLPGLRAP 139

15 Orf78: 122  VFVTAGISRKVSYLRFIIMDGLAA 145
          +++ +GI+R+VSY+RF++D AA
DedA:  140  IYVSGITRRVSYVREVLIDFCAA 163
  
```

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF78 shows 93.8% identity over a 145aa overlap with an ORF (ORF78a) from strain A of *N.*

meningitidis:

```

20 orf78.pep      10      20      30      40      50      60
          MFALFLEAFFVEYGYAAVFFVLVICGFGVPIDPDLTLVTGGVISGMGYTNPHIMFAVGMGLG
          |||:|||||
25 orf78a        10      20      30      40      50      60
          MFALFLEAFFVEYGYAAVFFVLVICGFGVPIDPDLTLVTGGVISGMGYTNPHIMFAVGMGLG
          |||:|||||

30 orf78.pep      70      80      90     100     110     120
          VLVDGIMFAAGRIWGQXXLFXPIAXIMTPKRYEQVQEKFDKYGNNWVLFVARFLPGLRT
          |||:|||||
35 orf78a        70      80      90     100     110     120
          VLVDGIMFAAGRIWGQKILKFKPIARIMTPKRYAQVQEKFDKYGNNWVLFVARFLPGLRT
          |||:|||||

          130      140
orf78.pep      AVFVTAGISRKVSYLRFIIMDGLAA
          |||:|||||
orf78a        AVFVTAGISRKVSYLRFIIMDGLAALISVFPVWYLGEYGAHNIDWIMAKMHSLSQSGIFTA
          130      140      150      160      170      180
  
```

The complete length ORF78a nucleotide sequence <SEQ ID 723> is:

```

40 1 ATGTTTGCC TTTTGAAGC CTTTTTGTG GAATACGGCT ATGCGGCCGT
51 GTTTTTCCT TGGTCATCT CGCGTTCGG COTCCGATT OCCGAGGATT
101 TGACCTTGGT AACAGCGCGC GTGATTTCGG GTATGGGPTA TACCAATCCG
151 CATATTATGT TTGCAGTCGG TATGCTCGG GTATTGGTCG GGGACGGCAT
201 CATGTTCCGC GCCGAGCACA TCTGGGGCCA GAAATCCTC AAGTTCARAC
251 CGATTGCGCG CATCATGACG CCGAAACGTT ACGCACAGST TCAGGAAAAA
301 TTCGACAAAT ACGGCACACT GGTGTTATTT GTCGCTCGTT TCCTGCCCGG
351 TTTGCGGACT GCGCTTTTCG TTACGCGCGG CATCAGCCGC AAGATATCGT
401 ATCTGCGCCT TCTGATTATG GACGGGCTTG CCGCGCTGAT TTCCGTGCC
451 GTTTGGATTG ACTTGGGCGA GTACGGCGCG CACAAATCGC ATTGGCTGAT
501 GCGCAAAATG CACAGCCTCG AATCGGCGAT CTTATCGCA TTGGGCGTGC
551 TGGGCGCGCG GCTGGCGTGG TTCTGTTGGC CAAAACGCGC ACATTATCAG
601 CTTTACCGCG CACAATTGAG CGAAAACGCG GCCAAACGCA AGGCGGAAAA
651 GGCACGCAAA AAGCGCGCAC AGAAGCAGCA GTAA
  
```

This encodes a protein having amino acid sequence <SEQ ID 724>:

```

55 1 MFALFLEAFF EYGYAAVFFV LVICGFGVPI PDLTLVTGG VISGMGYTNP
          HIMEAVGMGL VLVDGIMFAA AGRIWGQKIL KFKPIARIMT PKRYAQVQEK
101 FDKYGNWVLF VARFLPGLRT AVFVTAGISR KVSYLRFIM DGLAALISVP
  
```


151 VWIYLGEYGA HNIIDWLMKRN HSLQSGIFIA LGVLAALAW FWWKRRRHYQ
201 LYRAQLSEKR AKRKAERKAA KAAQKQ*

ORF78a and ORF78-1 show 89.0% identity in 227 aa overlap:

		10	20	30	40	50	60
5	orf78a.pep	MPALLERAFVEYGYAAVFFVLVICGFGVPIEDLTLTVTGGVISGMGTNPIMFAVGMLG					
	orf78-1	MPALLERAFVEYGYAAVFFVLVICGFGVPIEDLTLTVTGGVISGMGTNPIMFAVGMMLG					
		10	20	30	40	50	60
10	orf78a.pep	VLVGDGIMFAAGRIWQKILKFKPIARIMTPKRYAQVQVEKFDKYNWVLFVARFLPGLRT					
	orf78-1	VLVGDGIMFAAGRIWQKILKFKPIARIMTPKRYEQVQVEKFDKYNWVLFVARFLPGLRT					
		70	80	90	100	110	120
15	orf78a.pep	AVFVTAGISRKVSYLRFIIMDGLAALISVPMIYLGEYGAHNIDWLMKMHSLQSGIFIA					
	orf78-1	AVFVTAGISRKVSYLRFIIMDGLAALISVPMIYLGEYGAHNIDWLMKMHSLQSGIFVI					
		130	140	150	160	170	180
20	orf78a.pep	LGVLAAALAWFWWRKRHHYQLYRAQLSEKRAKRAKAAKAAQKQX					
	orf78-1	LGIGATVVAWVWKKRQRIQFYRSKLKEKRAKRAKAAKAAQKQX					
		190	200	210	220		
25	orf78a.pep	LGVLAAALAWFWWRKRHHYQLYRAQLSEKRAKRAKAAKAAQKQX					
	orf78-1	LGIGATVVAWVWKKRQRIQFYRSKLKEKRAKRAKAAKAAQKQX					
		190	200	210	220		

Homology with a predicted ORF from *N.gonorrhoeae*

ORF78 shows 97.4% identity over 38 aa overlap with a predicted ORF (ORF78ng) from *N.*

30	<i>gonorrhoeae</i> :		
	orf78.pep	XXLFXPIAXIMTPEXRYEQVEKFDKYNWVLFVARFLPGLRTAVFVTAGISRKVSYLRF	137
	orf78ng	YPVLFVARFLPGLRTAVFVTAGISRKVSYLRF	32
35	orf78.pep	IIMDGLAA	145
	orf78ng	LIMDGLAALISVPMIYLGEYGAHNIDWLMKMHSLQSGIFIALGVLAALAWFWWRKR	92

The ORF78ng nucleotide sequence <SEQ ID 72> is predicted to encode a protein comprising amino acid sequence <SEQ ID 726>:

40	1	..YPVLFVARFL PGLRTAVFVT AGISRKVSYL RFLIMDGLAA LISVPMIYL
	51	GEYGAHNIDW LMAKMHSLQS GIPALGVLA AALAWFWWRK RRHYQLYRAQ
	101	LSEKRAKRAK EKAAKAAQK QQ*

Further work revealed the complete gonococcal nucleotide sequence <SEQ ID 727>:

45	1	atgtttgccc tttTggaagc CTTTTTTGTC GAATaaggCt atgcGGCCGT
	51	GTTTTTCGTT TTGGTCATCT GCGGTTTCGG CGTSCCGATT CCCGAAGATT
	101	TGAACTTCGT AACGGCGCG GTGATTTGCG GTATGCGTTA TACCAATCCG
	151	CATATTATGT TTGCGGTCGG TATGCTCGCG GTGTTGGCGG GCGACGCGT
	201	GATGTTTGCC GCCGGAACGCA TCTGGGGCA GAAATCCTC AAGTTCAAAC
	251	CGATTGCGCG CATCATGACG CCGAAACGTT ACGGCGAGGT TCAGGAAAAA
50	301	TTCCAGCAAT ACGGCAACTG GGTTCGTGTT GTCGCCGCTT TCCTGCCGGG
	351	TTTGCGGACT GCGGTTTTCG TTACGCCGCG CATCAGCCGC AAGATATCGT
	401	ATCTGCGCTT TCTGATTATG GACGGGCTGG CCGCGCTGAT TTCCGTGCC
	451	GTTTGGATTG ACTTGGGCGA GTACGGCGCG CACAACATCG ATTTGGCTGAT
	501	GGCGAAATG CACAGCCTGC AATCGGCGAT CTTCATGCA TTGGGCGTGC
55	551	TGGCGGCGCG GCTGGCGTGG TTCTGTGGTG GCAACGCGC ACATTATCAG
	601	CTTTACGCGC CACAATTGAG CGAAAAACGC GCCAAACGCA AGGCGGAAAA
	651	GGCAGCGAAA AAGCGGCAC AGAAGCAGCA GTaA

This corresponds to the amino acid sequence <SEQ ID 728; ORF78ng-1>:

```

5      1 MFALLEAFFV EYGYAAVFFV LVICGFGVPI PEDLTLVTGG VISGMGYTNP
      51 HMFVAVGMLG VLAGDGMVFA AGRIWQGKIL KFKPIARIMT PKRYAQVQEK
      101 FDKYGNWVLF VARFLEPLRT AVFVTAGISR KVSYLRFILM DGLAALISVE
      151 WVIYLGEYGA HNDWLMAKM HSLQSGIFIA LGVLAALAW FWRKRRHYQ
      201 LYRAQLSEKR AKRKAQKAAK KAAQKQK*

```

ORF78ng-1 and ORF78-1 show 88.1% identity in 227 aa overlap:

```

10      10      20      30      40      50      60
      orf78-1.pep MFALLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGMGYTNP HMFVAVGMLG
      orf78ng-1 MFALLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGMGYTNP HMFVAVGMLG
      10      20      30      40      50      60

      70      80      90      100     110     120
      orf78-1.pep VLVDGDMFAAGRIWQKILKFKPIARIMTPKRYEQVQEKFDKYGNWVLFVARFLEPLRT
      orf78ng-1 VLAGDGMVFAAGRIWQKILKFKPIARIMTPKRYAQVQEKFDKYGNWVLFVARFLEPLRT
      70      80      90      100     110     120

      130     140     150     160     170     180
      orf78-1.pep AVFVTAGISRKVSYLRFILMDGLAALISVPWIYLGEGAHNIDWLMAKMHSLQSGIFVI
      orf78ng-1 AVFVTAGISRKVSYLRFILMDGLAALISVPWVIYLGEGAHNIDWLMAKMHSLQSGIFIA
      130     140     150     160     170     180

      190     200     210     220
      orf78-1.pep LGIGATVVAWVWKKRQRIQFYRSKLKEKRAQKAAKAAQSQKX
      orf78ng-1 LGVLAALAWFWRKRRHYQLYRAQLSEKRAKRAKAAKAAQKQKX
      190     200     210     220

```

Furthermore, orf78ng-1 shows homology to the dedA protein from *H. influenzae*:

```

35      sp|P45280|YG29 HAEM HYPOTHETICAL PROTEIN HI1629 >gi|1073983|pir||D64133 dedA
      protein (dedA) homolog - Haemophilus influenzae (strain Rd KW20)
      >gi|1574476 (U32836) dedA protein (dedA) [Haemophilus influenzae] Length = 212
      Score = 223 bits (563), Expect = 7e-58
      Identities = 108/182 (59%), Positives = 140/182 (76%), Gaps = 2/182 (1%)

      Query: 5 LEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGCGVISGM--GYTNP HMFVAVGMLGV 62
      L FF EYGY AV FVLICGFGVPIPED+TLV+GCVI+G+ N H+M V M+GVL
      40      Sbjct: 21 LIGFFTEYGYWAVLFVLLICGFGVPIPEDITLVSGGVITAGLYPENNVSHMLLVSMIGVL 80

      Query: 63 AGDGMVFAAGRIWQKILKFKPIARIMTPKRYAQVQEKFDKYGNWVLFVARFLEPLRTAV 122
      AGD M+ GRI+G KIL+F+PI RI+T +R V+EKF +YGN VLFVARFLEPLR +
      Sbjct: 81 AGDSCHYWLGRITYGTKILRFPIRIRIVTLQRLRMVREKFSQYGNRVLFVARFLEPLRAPI 140

      Query: 123 FVTAGISRKVSYLRFILMDGLAALISVPWVIYLGEGAHNIDWLMAKMHSLQSGIFIALG 182
      ++ +GI+R+VS+Y+R+T+D AA+ISVP+WVIYLG GA N+DWL ++ Q I+I +G
      Sbjct: 141 YNVSIGITRRVSVYRVLIDFCAALISVPWIYILGELGAKNLDWLHTQIQKGQIVVIYIFIG 200

      Query: 183 VL 184
      L
      Sbjct: 201 YL 202

```

Based on this analysis, including the presence of putative transmembrane domains, it is predicted that these proteins from *N. meningitidis* and *N. gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 87

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 729>:

5

10

```

1  ATCGAAAAAT  TTATGGCGGC  CGTGATGATG  GCAGGTTTGG  CAGCGCGCGT
51  TTGCAACGAC  GGAATCCAGC  TTGAGACAGC  CTGGGCGTGG  ACCACGATCG
101  AAGGTATGAA  AATAGACGCG  CGCTTCATGA  CCAATTCACG  CAGCAGAAGC
151  AAAACAAGAT  TTATGCTCGG  CGGAAAGCAG  CCGGTTCGCG  ACCGCTTCGA
201  AGTGCAGTCC  CACATCAACA  CGCAAGCGGT  TATCGGGATG  CGCAGAAGTG
251  AAGCGCGAGC  GCCTTTGGA  ACGAATCCGT  TTACGGAAGT  CAGACCGCG
301  AGCTATATAT  GGGTATATAT  GGGTATATAT  GGGTATATAT  GGGTATATAT
351  TAAATTTCTC  GTTACCTTGC  AATTTAAAAA  CGCAGAAAGC  CAAACGCTCC
401  AACTGAGAGT  GAAATATCGC  CGATGTGCGG  CAATGAACCA  C...

```

This corresponds to the amino acid sequence <SEQ ID 730; ORF79>:

1 MKKLLAAVMM AGLAGAVSAA GVHVEDGWAR TTVEGMKIGG AFMKIHNDAA
51 KQDFLLGGSS PVADRVVHT HINDNGVMRM REVEGGVPLE AKSVTELKPG
15 101 SYHVMFMGLK KOLKEGDKIP VTLKEFNAKA QTVOLEVKIA PMPAMNH..

Further work revealed the complete nucleotide sequence <SEQ ID 731>:

20 1 ATGAAAAAAT ATTTTGGCGCG CGTGAATGAG CGAGGTPTGG CAGGCGCGGT
51 TTCCGCGCCCG GAGTCCGCGC TTGAGGAOAG CTGGGCGCGG ACCACCGTGC
101 AAGGTATCGAA AATAGAGCGCG CGGTTCAATGA CAAATCACAA CGACGAAGCG
151 AAAACAAGACT TTGTGCTCGG ACCGAGCAGC CGGCTTGCAG AATCGCTCGA
201 AGTCGATACG CACATCAACG ACAACGCGGT GATCGCGGAT CGCGAAGTGC
251 AAGCGAGCGGT CCCTTTGGAA CGGAAATCGT TTACCGAGCT CAAACCGCGC
301 AGCTATCATG TGAATGTTAT GGGTTTGA AAACAATTTAA AAGAGGCGGA
351 TAAATAFTCC GTTACCTCGA AGTTTAAAAA CGCCAAAGCG CAAACCGTGC
25 401 AACTGGGAAGT CAAATCTCGG CGGATTCGCG CAAATGAACCA CGGTCACTCA
451 CACGGCGAAG CGCATCAGCA CTAA

This corresponds to the amino acid sequence <SEQ ID 732; ORF79-1>:

30

```
1 MKKLLAAVMM AGLAGAVSAA GVHVEDGWAR TTVEGMKIGG AFMKIHNDDEA
51 KQDFLLGGSS PVADRVVEHT HINDNGVMRM REVEGGVPLE AKSVTELKPG
101 SYHVMFMGLK KQLKEGDKIP VTLKFNKA QTVQLEVKIA PMPAMNHGHH
151 HGEAHOH*
```

Computer analysis of this amino acid sequence revealed a putative leader peptide and also gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

35 ORF79 shows 94.6% identity over a 147aa overlap with an ORF (ORF79a) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
	orf79.pep	<u>MKKLLAAVVMAGLAGAVSAAGVHVEDG</u> WARTTVEGMMKIGGAFMKHNDNEAKQDFLLGGSS					
40	orf79a	<u>MKKLLAAVVMAGLAGAVSAAGIHVEDG</u> WARTTVEGMMKGGGAFMKHNDNEAKQDFLLGGSS					
		10	20	30	40	50	60
	orf79.pep	PVADRVVEVHTHINDGVMRMREVEGGVPLEAKSVTELPFGSYHVMFMGLKKQLKEGD	120				
45	orf79a	PVADRVVEVHTHINDGVMRMREVEGGVPLEAKSVTELPFGSYHVMFMGXKQLKXGDK	120				
		70	80	90	100	110	120
	orf79.pep	VTLEKFNKAKGTQVLEVKIAPMPMANH					
	orf79a	VTLEKFNKAKGTQVLEVKIAPMSAMDGHIIHGEAHQH					
		130	140	150			
50	orf79.pep	VTLEKFNKAKGTQVLEVKIAPMPMANH					
	orf79a	VTLEKFNKAKGTQVLEVKIAPMSAMDGHIIHGEAHQH					
		130	140	150			

The complete length ORF79a nucleotide sequence <SEQ ID 733> is:

```

1   ATGAANAAC  TATTGGCAGC  CGTGATGATG  GCAGGTTTGG  CAGGCGCGGT
51  TTCCGCCGCC  GGAATCCACG  TTGAGGACGG  CTGGGCGCGC  ACCACCGCTG
101 AAGGTATGAA  AATGGCGGCG  GCGTTCATGA  AAATCCACAA  CGACGAAGCC
151 AAACAAGACT  TTTTGCTCGG  CGGAAGCAGC  CCTGTGCGCG  ACCGCGTGA
201 AGTGCAATAC  CATATCAATG  ATAACGGTGT  GATGCGGATG  CGCGAAGTCG
251 AAGGCGGCGT  GCCTTTGGAG  GCGAAATCG  TTACCGAATC  CAAACCCGCG
301 AGCTATCATG  TCATGTTTAT  GGGTNTGAAA  AAACAATTAA  AAGANGGGGA
351 CAAGATTCCC  GTTACCTGTA  AATTTAAAAA  CGCCAAAGCA  CAAACCGTCC
401 AACTGGAGT  CAAAACGCG  CCGATGTCGG  CAATGGACCA  CGGTATCATC
451 CACGGGGAAG  CGCATCAGCA  CTA

```

This encodes a protein having amino acid sequence <SEQ ID 734>:

```

1   MKXLLAAVMM  AGLAGAVSAA  GIHVEDGWAR  TTVEGMMGG  AFMKIHND
151 KQDFLLGGSS  FVADRVEVHT  HINDNGVMRM  REVEGGVPLE  AKSVTELKPG
101 SYHVMFMGXK  KQLKXGDKIP  VTLKFNKAQA  QTVQLEVKTA  PMSAMDHGHH
151 HGEAHQH*

```

ORF79a and ORF79-1 show 94.9% identity in 157 aa overlap:

		10	20	30	40	50	60
20	orf79a.pep	MKXLLAAVMMAGLAGAVSAAGI	HVEDGWARTTVEGMMGGAFMKIHND	EAKQDFLLGGSS			
	orf79-1	MKKLLAAVMMAGLAGAVSAAGV	HVEDGWARTTVEGMMGGAFMKIHND	EAKQDFLLGGSS			
		10	20	30	40	50	60
25	orf79a.pep	FVADRVEVHTHINDNGVMRM	REVEGGVPLEAKSVTELPKPSYHVMFMGXK	QQLKXGDKIP			
	orf79-1	FVADRVEVHTHINDNGVMRM	REVEGGVPLEAKSVTELPKPSYHVMFMGLKK	QQLKEGDKIP			
		70	80	90	100	110	120
30	orf79a.pep	VTLKFNKAQAQTVQLEVKTA	PMSAMDHGHHHGEAHQH				
	orf79-1	VTLKFNKAQAQTVQLEVKTA	PMPAMNHGHHHGEAHQH				
		130	140	150			
35	orf79a.pep	VTLKFNKAQAQTVQLEVKTA	PMPAMNHGHHHGEAHQH				
	orf79-1	VTLKFNKAQAQTVQLEVKTA	PMPAMNHGHHHGEAHQH				
		130	140	150			

Homology with a predicted ORF from *N.gonorrhoeae*

ORF79 shows 96.1% identity over 76 aa overlap with a predicted ORF (ORF79ng) from *N.gonorrhoeae*:

```

40  orf79.pep      FMKIHND EAKQDFLLGGSSFVADRVEVHTHINDNGVMRMREVEGGVPLEAKSVTELPKFS 101
      orf79ng      INDNNGVMRMREVEGGVPLEAKSVTELPKFS 30
45  orf79.pep      YHVMFMGLKKQLKEGDKIPVTLKFNKAQAQTVQLEVKTA PMPAMNH 147
      orf79ng      YHVMFMGLKKQLKEGDKIPVTLKFNKAQAQTVQLEVKTA PMSAMNHGHHHGEAHQH 86

```

An ORF79ng nucleotide sequence <SEQ ID 735> was predicted to encode a protein comprising amino acid sequence <SEQ ID 736>:

```

1   ..INDNGVMRM  EVKGGVPLEA  KSVTELPKFS  YHVMFMGLKK  QLKEGDKIPV
51  TLKFNKAQAQ  TVQLEVKTA  PMSAMNHGHH  HGEAHQH*

```

Further work revealed the complete gonococcal DNA sequence <SEQ ID 737>:

```

1   ATGAAAAAAT  TATTGGCAGC  CGTGATGATG  GCAGGTTTGG  CAGGCGCGGT
51  TTccgccgcC  GgagTccAtG  TCGAggACGG  CTGGCGCGCG  accaCTgtcg
101 aaggtATgaa  aatggcGCGG  GcgttCATgA  aaATCCACAA  CGACGaaGcc
151 atacaaGACT  ttgtgcTCgg  CGGaagcatg  ccggttcgcy  accgcGTCGA
201 AGTGCAcata  cacATCAACG  ACAACGGCGT  GATGCGTATG  CGCGAAGTCA

```

251 AAGGCGGCGT GCCTTTGGAG GCGAARTCGG TTACCGAACT CAAACCCGGC
 301 AGCTATCACG TGATGTTTAT GGGTTTGAAA AAACAACCTGA AAGAGGGCGA
 351 CAAGATTCCC GTTACCCTGA AATTTAAAAA CGCCAAAGCG CAACACCGTCC
 401 AACTGGAAGT CAAACCCGCG CCGATGTGGG CAATGAACCA CCGTCATCAC
 451 CACGCGGAAG CGCATCAGCA CTAA

This corresponds to the amino acid sequence <SEQ ID 738; ORF79ng-1>:

1 MKKLLAAVMM AGLAGAVSAA GVHVEDGWAR TTVEGMKMG AFMKIHND
 51 IQDFVLGGSM PVADRVVHT HINDNGVMRM REVKGVPLE AKSVTELKPG
 101 SYHVMFMGLK KQLKEGDKIP VTLKFKNAKA QTVQLEVKA PMSAMNHGH
 151 HGEAHQH*

ORF79ng-1 and ORF79-1 show 95.5% identity in 157 aa overlap:

		10	20	30	40	50	60
orf79-1.pep		MKKLLAAVMMAGLAGAVS	AAGVHVEDG	WARTTVEGMKMG	AFMKIHND	EAKQDFLLGGSS	
orf79ng-1		MKKLLAAVMMAGLAGAVS	AAGVHVEDG	WARTTVEGMKMG	AFMKIHND	EAIQDFVLGGSM	
		10	20	30	40	50	60
		70	80	90	100	110	120
orf79-1.pep		PVADRVVHTHINDNGVMRM	REVKGVPLEAKSVTELKPG	SYHVMFMGLK	KQLKEGDKIP		
orf79ng-1		PVADRVVHTHINDNGVMRM	REVKGVPLEAKSVTELKPG	SYHVMFMGLK	KQLKEGDKIP		
		70	80	90	100	110	120
		130	140	150			
orf79-1.pep		VTLKFKNAKAQTVQLEV	KIAPMFAMNHGHHG	GEAHQH			
orf79ng-1		VTLKFKNAKAQTVQLEV	KIAPMFAMNHGHHG	GEAHQH			
		130	140	150			

Furthermore, ORF79ng-1 shows significant homology to a protein from *Aquifex aeolicus*:

gi|2983695 (AE000731) putative protein [Aquifex aeolicus] Length = 151
 Score = 63.6 bits (152), Expect = 6e-10
 Identities = 38/114 (33%), Positives = 58/114 (50%), Gaps = 1/114 (0%)

Query: 24 VEDGWARTTVEGMKMGAFMKIHND EAIQDFVLGGSPVADRVVHTHINDNGVMRMREV 83
 v+ w G M I N+ D+++G +A RVE+H + +N V +M

Sbjct: 27 VKHPVWMEPPPGPNTTMMGMIIIVNEGDEPDYLIQAKTDIAQRVELHKTVIENDVAKMVPQ 86

Query: 84 KGGVPLEAKSVTELKPGSYHVMFMGLKKQLKEGDKIPVTLKFKNAKAQTVQLEV 137
 + + + K E K YHVM +GLKK++KEGDK+ V L F+ + TV+ V

Sbjct: 87 ER-IEIPPKGKVEFKHHGYHVMIIIGLKKRIKEGDKVKVELIFEKSGKITVEAPV 139

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF79-1 (15.6kDa) was cloned in the pET vector and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 18A shows the results of affinity purification of the His-fusion protein. Purified His-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result) and FACS analysis (Figure 18B). These experiments confirm that ORF79-1 is a surface-exposed protein, and that it is a useful immunogen.

Example 88

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 739>:

5	1	ATGACGGTAA	CTGGCGCGCA	AGGCGGCAAA	GCTGCCAAGG	CGTTAAAAAA
	5	ATATCTGCTT	ACGGGCAATT	TGGTCTCGCTT	CGCGATTGGG	ATACACGTTT
	101	GGGTGGTCTT	CTATATCTGT	TCCCGCTGGT	ATCAGCTTCG	CACCTCGTGT
	151	CGAGCAACAT	GGGCGCGCGA	ATAATGTTTG	GGGTTTAATA	TCCCGGGGCT
	201	GGGCGTTTAT	CTGTGCGATG	ACGCTATTGT	TGTAAACCGA	TGTTTTCGGC
10	251	CCAACGTATT	GGGTCCGACG	CCGTCCTGGC	CTGGGACCGA	CCTGTTGGCG
	301	CGGATTTCGG	TGTGGAATCT	CATCTATCTG	AGGTGTAAAA	AAGTATCCGA
	351	ATAcgTGCCT	TCCGACAGCA	CGCGTTCGTT	TAAAGACCGC	GTACTCGTGG
	401	CGCTTCCCCA	CGCGCGTATT	TGGACATGAT	CTTTCGCTGT	AGGGCGATTC
	451	TGCAATTCCG	TATAGGCGTC	ATGAGCCGA	GATCCGTCG	ATCTTCTGTT
15	501	TGATATTCCG	ACCGACCGGA	ATTCCGACGG	CGGTACATAT	ATAGCGTCTG
	551	AGTAAAGGCA	TCTGACGAGA	CTGTCGACCA	CGCTGGACGA	AsCATTTGAA
	601	TATCTGCTAT	CGCTGGTATT	GGTCACTCC	GACAGCACTG	CGCTCAAAAC
	651	ATTGGCAsGA	CCTATGCCGT	CTGAAAAGCG	GGATTTCGCC	GAACACAAT
	701	AA				

This corresponds to the amino acid sequence <SEQ ID 740; ORF98>:

20	1	MTVTAAREGGK	AAKALKKYLI	TGILVWLPIA	VTVWVVSIVY	SASDQLVNL
	51	PKQWAPQYVL	GFNIPGLGVI	VATAVLVFTG	LFAANVLGRQ	ILAAWSDSLG
	101	IRIPVVKSIYS	GVNKSEYVL	SDSSRSFKTP	VLVPPQPQGI	WTIAFVSQGI
	151	SNVAKAALPL	GDGLVSVYVP	TTPNPTGGYI	IMVKSKDVRE	LDMSVDXLK
	201	YVAILGSMVIP	DDLEPVKTAX	EMPEKKAIDL	EQQ*	

25 Further work revealed the complete nucleotide sequence <SEQ ID 741>:

	1	ATACCGGAAAC	nTGGCGGCCGA	AGCCCGCCGAA	TGGTCCGAAC	CGTTAAAAAA
	5	ATATCTGATT	AGCGCGATT	TGGTCTCGAT	CGCGATTGCG	GTACCGGTTT
	10	GGGTGGTGTTC	CTATATCTGTT	TCCCGCTTGG	ATCAGCTCGT	CAACCTGCTG
30	151	CGCAAGCAAT	GGCGGCCGCA	ATATGTTTTCG	GGGTTTAATA	TCCCGGGCTG
	201	GGCGCTTATC	TGTGCGATTG	CGGATATGTT	TGTAAACCGA	TTGTTTGCGC
	251	CCAACTGATT	GGGTGGCGAG	ATCTCGCCGC	CGTGGACAG	CTGTGTCGGA
	301	CGGATTCGGT	TGTGAAATCT	CATCTATTGC	TAGTGTAAAA	AAGTATCCGA
	351	ATCTCGTGTG	TCCGACAGCA	CGCGTTCGTT	TTAAACGCGC	GTACTCTGTG
35	401	TCGTTTCCCC	CGCCCGGTTT	TGGACAGATT	CTTCCTGTCT	AGGCGAGGTT
	451	CGTAATGGCT	TTAGGCGGAT	ATTGCGCAAT	GATCTTCTGC	ATCTTCTGCT
	501	GTAATGTCGG	CGTGGCGGAT	ATCCGAGAT	TTATATGGCT	TTATATGGCT
	551	AGAAATGAT	TCCGTCGGAA	CTCGATATCA	CGCTGCGACA	AGCATTTGAA
	601	TATCTGATT	CGCTGGGTAT	GGTCACTCTC	GACGACCTGC	CGCTCAAAC
	651	ATTGGCAGGA	CCTATGCCGT	CTGAAAAGCG	GGATTTGCC	GAACAAACAT
40	701	AA				

This corresponds to the amino acid sequence <SEQ ID 742; ORF98-1>:

45

1	MTEXAEEGGK	AAKALKKYL	TGILVWLPIA	VTVWVVSIV	SASDQLVNL
51	PQKWRPQVYL	GNFIPGLGVI	VATAVLFTG	LFAANVLGRQ	ILAAWDSLIG
101	RIPVVKSIYS	SVKKSESEL	SDSRSRFTK	VLVPPPGQGI	WTISAFVSGO
151	SNAVKAALPK	DDGLVSVYP	TNPNTGADY	IMVKSKDVRE	LDMSVDEAIK
201	YVISGMVIP	DGPEVKTLAG	EMPESEKKPL	EQQ*	

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF98 shows 96.1% identity over a 233aa overlap with an ORF (ORF98a) from strain A of *N.*

50 meningitidis:

orf98.pep 10 20 30 40 50 60
MTVTAAEGGKAAKALKKYLITGILVWLPIAVTVWVSYIVSASDQLVNLLPKQWRPQYVL

5	orf98a	MTEPAAEGGKAALKKKYLITGILVWLPIAVTVVWVSIVSASDQLVNLLEPKQWRPQYVL	10	20	30	40	50	60
	orf98.pep	GFNIPGLGVIAIAVLFTGLFAANVLGRQILAAWDSLGRIPVVKSIYSSVKKVSXSVLL	70	80	90	100	110	120
	orf98a	GFNIPGLGVIAIAVLFTGLFAANVLGRQILAAWDSLGRIPVVKSIYSSVKKVSXSVLL	70	80	90	100	110	120
10	orf98.pep	SDSSRSFKTPVLVFPFQSGIWTIAFVSGQVSNVAKAALPKDGDYLSVYVPTTNPNTGGYY	130	140	150	160	170	180
	orf98a	SDSSRSFKTPVLVFPFQSGIWTIAFVSGQVSNVAKAALPKDGDYLSVYVPTTNPNTGGYY	130	140	150	160	170	180
	orf98a	SDSSRSFKTPVLVFPFQSGIWTIAFVSGQVSNVAKAALPKDGDYLSVYVPTTNPNTGGYY	130	140	150	160	170	180
15	orf98.pep	IMVKKSDVRELDMSVDEALKYVISLGMVIPDDLFPVKTLAGPMPSEKADLPEQQX	190	200	210	220	230	
	orf98a	IMVKKSDVRELDMSVDEALKYVISLGMVIPDDLFPVKTLAGPMPSEKADLPEQQX	190	200	210	220	230	
	orf98a	IMVKKSDVRELDMSVDEALKYVISLGMVIPDDLFPVKTLAGPMPSEKADLPEQQX	190	200	210	220	230	

The complete length ORF98a nucleotide sequence <SEQ ID 743> is:

25	1	ATGACGGAAC	CTGCGGCCGA	AGGCGGCAAA	GCTGCCAAGG	CGTTAAAAAA
	51	ATATCTGATT	ACGGGCATT	TGGTCTGGCT	GCCGATTGCG	GTAACGGTTT
	101	GGGTGGTTTC	CTATATCGTT	TCCGCGTCOG	ATCAGCTCGT	CAACCTGCTG
30	151	CCGAAGCAAT	GGCGGCCGCA	ATATGTTTTG	GGGTTTAATA	TCCCGGGGCT
	201	GGCGGTTATC	GTTGCCATTG	CCGTATTGTT	TGTAACCCGA	TTATTTGCCG
	251	CAAACTGATT	GGCGCGGCAG	ATTCITGGCC	CGTGGGACAG	CTTGTGGGG
35	301	CGGATTCGGG	TGTGAAGTTC	CATCTATTCC	AGTGTGAAA	AAGTATCCGA
	351	NTCGTTGCTG	TCCGACAGCA	GCCGTTGCTT	TAAACACCCA	GTACTCGTGC
	401	CGTTTCCGCA	ATCGGGTATT	TGGACARTCG	CATTCGTGTC	CGGTACAGTG
40	451	TGCAATGGCG	TTAAGCGCGC	ATTCGCGAAG	GACGCGGATT	ATCTTCCGCT
	501	GTATGTTCCG	ACCAACGCGA	ATCCGACCGG	CGCTGCTAAT	ATATATGAAA
	551	AGAAAAAGCA	TGTGCGCGAA	CTCGATATGA	CGCTGGACGA	ACGCTTGAAA
45	601	TATGTGATT	CGCTGGGTAT	GGTCATCCCT	GACGACCTGC	CCGTCAAAAC
	651	ATTGGCAGGA	CCTATGCCGT	CTGAAAAGGC	GGATTGGCCC	GAACAACAA
	701	AA				

This encodes a protein having amino acid sequence <SEQ ID 744>:

40	1	MTEPAAEGGK	AAKALKKYL	ITGILVWLPI	AVTVVWVSIV	SASDQLVNL
	51	PKQWRPQYVL	GFNIPGLGVI	VAIAVLEVTG	LFRANVLGRQ	ILAAWDSL
	101	RIPVVKSIYS	SVKKVXSLL	SDSSRSFKTP	VLVFPFQSGI	WTIAFVSGQV
45	151	SNVAKAALPK	DGYLSVYVP	TTPNTGGYY	IMVKKSDVRE	LDMVDEALK
	201	YVISLGMVIP	DDLFPVKTLAG	PMPSEKADLP	EQQ*	

ORF98a and ORF98-1 show 98.7% identity in 233 aa overlap:

45	orf98a.pep	MTEPAAEGGKAALKKKYLITGILVWLPIAVTVVWVSIVSASDQLVNLLEPKQWRPQYVL	10	20	30	40	50	60
	orf98-1	MTEPAAEGGKAALKKKYLITGILVWLPIAVTVVWVSIVSASDQLVNLLEPKQWRPQYVL	10	20	30	40	50	60
	orf98a.pep	GFNIPGLGVIAIAVLFTGLFAANVLGRQILAAWDSLGRIPVVKSIYSSVKKVSXSVLL	70	80	90	100	110	120
50	orf98a.pep	GFNIPGLGVIAIAVLFTGLFAANVLGRQILAAWDSLGRIPVVKSIYSSVKKVSXSVLL	70	80	90	100	110	120
	orf98-1	GFNIPGLGVIAIAVLFTGLFAANVLGRQILAAWDSLGRIPVVKSIYSSVKKVSXSVLL	70	80	90	100	110	120
	orf98a.pep	SDSSRSFKTPVLVFPFQSGIWTIAFVSGQVSNVAKAALPKDGDYLSVYVPTTNPNTGGYY	130	140	150	160	170	180
55	orf98-1	SDSSRSFKTPVLVFPFQSGIWTIAFVSGQVSNVAKAALPKDGDYLSVYVPTTNPNTGGYY	130	140	150	160	170	180
	orf98a.pep	IMVKKSDVRELDMSVDEALKYVISLGMVIPDDLFPVKTLAGPMPSEKADLPEQQX	190	200	210	220	230	
	orf98-1	IMVKKSDVRELDMSVDEALKYVISLGMVIPDDLFPVKTLAGPMPSEKADLPEQQX	190	200	210	220	230	

orf98-1 IMVKKSDVRELDMSVDEALKYVISLGMVIPDDLPVKTLAGMPSEKADLPQQX
 190 200 210 220 230

5 Homology with a predicted ORF from *N.gonorrhoeae*

ORF98 shows 95.3% identity over a 233 aa overlap with a predicted ORF (ORF98ng) from *N.gonorrhoeae*:

		10	20	30	40	50	60	
10	orf98.pep	MTVTA	AEGGKA	AKALKK	YILTGIL	VWLPIA	IVTVV	VSVIVS
	orf98ng	MTEPA	AEGGKA	AKALKK	YILTGIL	VWLPIA	IVTVV	VSVIVS
	orf98.pep	GFNIP	GLGVIA	IVAVL	FVTGL	FAANVL	GRQILA	PAWDS
	orf98ng	GFNIP	GLGVIA	IVAVL	FVTGL	FAANVL	GRQILA	PAWDS
15	orf98.pep	SDSSR	SFKTPV	LVVFP	QSGI	WTIAF	VSGQV	SNVKA
	orf98ng	SDSSR	SFKTPV	LVVFP	QSGI	WTIAF	VSGQV	SNVKA
20	orf98.pep	IMVKK	SDVREL	DMSV	DEALKY	VISLGM	VIPDDLP	VKTLA
	orf98ng	IMVKK	SDVREL	DMSV	DEALKY	VISLGM	VIPDDLP	VKTLA

The complete length ORF98ng nucleotide sequence <SEQ ID 745> is predicted to encode a protein

25 having amino acid sequence <SEQ ID 746>:

30	1	MTEPA	AEGGK	AAKALK	KYLI	TGILV	WLPIA	IVTVV	VSVIV	SASD	QLVNLL
	51	PKQWR	PQYVL	GFNIP	GLGVI	VAIAV	LEVTG	LFAAN	VLGRQ	ILAAW	SDLLX
	101	RIPV	KSIYS	SVKKV	SESL	SDSSR	SFKTP	VLVFP	QSGI	WTIAF	VSGQV
	151	SNAVK	AALPQ	DGYLS	VYVP	TPNPT	GGYY	IMVKK	SDVRE	LDMSV	DEALK
	201	YVISL	GMVIP	DDL	FPVKTL	AG	MPPEK	ALP	EQQ*		

Further work revealed the complete nucleotide sequence <SEQ ID 747>:

35	1	ATGAC	GGAA	CTGCG	CCGA	AGCGG	CAAA	GCTGC	CAAG	CGTTA	AAAAA
	51	ATATC	TGATT	ACAGG	CATT	TGGT	CTGG	CGC	GATT	CGC	GTAAC
	101	GGGT	GGTTT	CTAT	ATCGT	TCCG	CGTCC	ACC	GCTT	TGT	CAACCT
	151	CCGA	AGCA	AT	ATGTTT	G	GGT	TTA	TA	TCCC	CGGG
	201	CGG	CGTAT	TTG	TCCAT	CG	TAT	TTG	T	GTAA	CCGGA
	251	CAAC	GTGTT	GGG	CGGCA	AT	CTT	CGC	CGT	GGC	AG
	301	cggat	TC	CGG	T	TTG	T	CA	AAT	C	AT
	351	ATC	GTG	CTG	T	TCC	GAC	G	CA	G	CCG
	401	CGTT	CCCCA	AT	CGG	TAT	TGG	ACA	AT	CG	CA
	451	T	CGA	AT	G	CGG	CGC	AT	T	G	CG
	501	GTAT	GT	CCG	ACC	ACG	CCG	CG	GT	T	ACT
	551	AGAAA	AG	CGA	TGT	CG	CGA	CT	C	AT	AT
	601	TAT	GT	GAT	TT	CG	T	GG	TAT	CG	TA
	651	ATT	GC	G	AG	CA	G	AT	CG	CG	CA
	701	AA									

This corresponds to the amino acid sequence <SEQ ID 748; ORF98ng-1>:

50	1	MTEPA	AEGGK	AAKALK	KYLI	TGILV	WLPIA	IVTVV	VSVIV	SASD	QLVNLL
	51	PKQWR	PQYVL	GFNIP	GLGVI	VAIAV	LEVTG	LFAAN	VLGRQ	ILAAW	SDLLG
	101	RIPV	KSIYS	SVKKV	SESL	SDSSR	SFKTP	VLVFP	QSGI	WTIAF	VSGQV
	151	SNAVK	AALPQ	DGYLS	VYVP	TPNPT	GGYY	IMVKK	SDVRE	LDMSV	DEALK
	201	YVISL	GMVIP	DDL	FPVKTL	AG	MPPEK	ALP	EQQ*		

ORF98ng-1 and ORF98-1 show 97.9% identity in 233 aa overlap:

55	orf98-1.pep	10	20	30	40	50	60
		MTEPA	AEGGKA	AKALKK	YILTGIL	VWLPIA	IVTVV

5	orf98ng-1	MTEPAEEGGKAAKALKKYLITGILVWLPIAVTVVWVSYIVSASDQLVNLPPKQWRPQYVL	10	20	30	40	50	60
			70	80	90	100	110	120
10	orf98-1.pep	GFNIPGLGVIAIAVLVFTGLFAANVLGRQILAAWDSLGRIPVVKSIYSSVKVSESL						
	orf98ng-1	GFNIPGLGVIAIAVLVFTGLFAANVLGRQILAAWDSLGRIPVVKSIYSSVKVSESL	70	80	90	100	110	120
15	orf98-1.pep	SDSSRSFKTPVLVPPFPQGIWTIAFVSGQVSNVAKALPKDGDYLSVYVPTTNPPTGGYY	130	140	150	160	170	180
	orf98ng-1	SDSSRSFKTPVLVPPFPQGIWTIAFVSGQVSNVAKALPKDGDYLSVYVPTTNPPTGGYY	130	140	150	160	170	180
20	orf98-1.pep	IMVKKSDVRELDMSVDALKYVISLGMVIPDDLPVKTLAGPMFESEKADLPEQQX	190	200	210	220	230	
	orf98ng-1	IMVKKSDVRELDMSVDALKYVISLGMVIPDDLPVKTLAGPMFESEKADLPEQQX	190	200	210	220	230	

Based on this analysis, including the fact that the putative transmembrane domains in the gonococcal protein are identical to the sequences in the meningococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 89

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 749>:

1	ATgAAAACGG	TAGTCTGGAT	TGTCGTCCTG	TTTGCCGCCG	CCGTCGGACT
51	GGCGCTGGCT	TCGGGCATTT	ACACCGGCGA	CGTGTATATC	GTACTCGGAC
101	AGACCATGCT	CAGAATCAAC	CTGCACGGCT	TTGTGTTAGG	TTCGCTGATT
151	GCGCTCGTGG	TGTGGTATTT	CTTGTTTAAA	TTCATTATCG	GgGdTACTCA
201	ATATCCCCGA	AAAGATGCAG	CGTTTCGGTT	CGGCGGTAA	AGGCCCKAAG
251	ssCGsGCTTG	CCTTGAACAA	GGCGGTTTGG	GCSTATTTTG	AAGGCGCTT
301	TGAARAAGCG	GAACTGAAAG	CTCAGCGGT	GTTCGTCAC	AAAGTAGGCC
351	GaGAGACAC	CGGACTTTGG	CATTGATGCT	GCGCGGCAC	GCGCGCGGAC
401	AGATGGAAAA	CATCGAsTGG	CGGACCGGTT	ATCTTCGGCA	AATCGCCAAA
451	CTGCCCGGAAA	AACAGCAGCT	TTCGGCTTAT	CTTTTGTGGT	CGGAATCGGCG
501	GTTGAACCGG	CGCGATTACG	AAGCGGCGGA	AGCCATATCT	CATGCGGCGG
551	CGAAGATGAA	TGCCAACCTT	ACGGCGCTCG	TGCGTCTGCA	.ATTGCTTAC
601	.GCTTTCGACA	GGGCGCAGCG	GTTGCAAGTT	CTGGCAAAAA	CCGAAAAAAT
651	TTCCAAGCGG	GGCGGCTTGG	GCAATCGGA	AATGGAACGG	TATCAAAATT
701	GGGCATATCC	GTCGCCAGCT	GGCGGATGCT	GCCGATGCCG	CCGCTTTGAA
751	AACCTGCCTG	AAGCGGATTC	CGACAGCCT	CAAAAACGGG	GAATTGAGCG
801	TATCGGTTCG	GGAAAAGTAC	GAACGTTTGG	GACTGTATGC	CGATGCGGCT
851	AAATGGGTCA	AACAGCATTAT	TCCGCAsAAC	CGCGGCCCGG	AGCTTTTGGG
901	AGCCTTTTGT	GAAAGGTCGC	GCTTTTGGGG	CGAGCGCGGA	CAGCAAAAAG
951	CCATCGATTT	TGCGGATGCT	TGGCTGAAG	AACAGCCGGA	TACGCGCTTT
1001	CTGCTGATGT	ATCTCGGTGG	CGCTCGCTTC	GCGCGCAAC	TTTGGGCAAC
1051	GGCAAAAGCG	TACCTTGAAG	CGACGATGCT	ATTAAGCCCG	AGTATTTCGG
1101	CGCTTTGCT	TCTAACAAG	GTTTTCGACG	AAATCGGAGA	ACCGCAGAAG
1151	GCGGAGCGCG	AC...			

This corresponds to the amino acid sequence <SEQ ID 750; ORF100>:

1	MKTVMVIVVL	FRAAVGLALA	SGIYTDGVYI	VLGQTMRLIN	LHAFVLGSLI
51	AVVVVYFLFK	FIIGVLNIPE	KMORFGSARK	GKXKXIALNK	AGLAYFFGFR
101	EKAELASERV	LVNKGVRDNR	TLALMLXAHA	AGQEMNIXRR	DRYLAEIAKL
151	PEKQQLSRYL	LAESALNRR	DYEAEEANLH	AAARMNANLT	RLVRLXIRYA
201	FDRGDALQVL	AKTEKLSKAG	ALGKSEMERY	QNWAYRRQLA	DAADAAALEK
251	CLKRIPIPSLK	NGELSVSVAE	KYERLGLYAD	AVKWVQKHP	XNRRPELLEA
301	FVESVRFLEGE	REQQKAIDFA	DAWLKEQPDN	ALLMLYLGLR	AFGRKLWGRA

351 KGYLEASIAL KPSISARLVL TKVFDEIGEP QKAEAH...

Further work revealed the complete nucleotide sequence <SEQ ID 751>:

```

1  ATGAAAACGG TAGCTGGAT TGTGCTCGT TTTGCGCCGG CGGTGGGACT
51  GCGCGCGGCT TCGGCAATT CGGTATATC GCGTATATC GCGTATATC
101 AGACCATGCT CAGATCAAC CTGCACGCT TGTGTTTAG TGTGTTGAT
151 GCGGTGCTGG TGTGTTATT CTGTTTAA TGCATTATC GCGTACTCA
201 TATCCCCGAA AAGATGCAG CTTTGGTTC GCGCGTAA GCGCGCAGG
251 CCGCGCTTGC CTGGAACAAG GCGGTTTGG CGTATTTGA AGGCGTTTT
301 GAAAAGCGGG AACTAGAAG CTCACGCGT TGTGTCACA AAGAGCGCG
351 AGACAACCGG ACTTTGGCAT TGATGCTGG GCGCGACGC GCGCGACAG
401 TGGAAAACAT CGAGCTGCG CAGCGTTAT TTTGCGGAA TGCCTAACT
451 CCGGAAAAC AGCAGCTTC CCGTTATCT TTTGTTGCG AATCGCGGT
501 GAACCGGCGC GATTACGAG CGGCGGAAG CAATCTTCA TCGGCGCGG
551 AGATGAATGC CAACCTTAC CGCCTCGTG GTCTGCACT TCGTTACGT
601 TTCGACAGGG GCGACGGGT GCAGGTTCT GCAAAAACG AAAAAGCTT
651 CAAGGCGGGC GCGTTGGGA AATCGGAA TGAACGGTA CAATTTGGG
701 CATACCGCG CAGCTGCGG GATGCTGCG ATGCGCGCG TTTGAAAAC
751 TGCCTGAAGC GGATTCCGA CAGCCTCAA AACGGGAA TGAAGGTAT
801 GGTTCGGGAA AAGTACGAC GTTTGGGACT GTATGCCAT CGGTCGAAT
851 GGCTCAACGA CATTATTCG CACAAACGC GCCCGAGCT TTTGGAGCG
901 TTTGTGAAA CGGTGCGCT TTTGGCGAG CCGAGACGC AGAAGCCAT
951 CGATTTTGCC GATGCTTGG TGAAGAACA CGCGATAA CGCTTCTGC
1001 TGATGTATCT CGGTGCGCT GCCTACGCC GCAACTTTG GCGCAAGCA
1051 AAAGGCTACC TTGAAGCGA CATTGCATT AAGCCAGTA TTTCCGCGC
1101 TTTGTTCTA GCAAGGTTT TCGACGAAT CGGAGAACG CAGAAAGCG
1151 AGGCGCAGG CAACTTGGT TTGAAGCG TCTCCGATG GAAAGCTCA
1201 GCAGCGTTAG AGCAGCATAG CTGA

```

This corresponds to the amino acid sequence <SEQ ID 752; ORF100-1>:

```

1  MKTVVWIVVL FAAAVGLALA SGIYTGDOYI VLQOTMLRN LHAFLVGLSLI
51  AVVWYFLFK FIIGVLNIPE KMRFGSARK GRKAAALNKK AGLAYFEGRF
101 EKAELEASRV LVNKEAGDNR TLALMLGAHA AQMENIELR DRYLAELAKL
151 PEKQQLSRYL LLAESALNRR DYEAAEANLH AAAKMNANLT RIVRLQLRYA
201 FDRGDALQVL AKTEKLSKAG ALCKSEMERV QNWAYNRQLA DAADAAALKT
251 CLKRLPDSLK NGELSVSVAE KYERLGLYAD ALLKWLQYH FHNRRFELLEA
301 FVESVRFLEGE REQQKAIDFA DAWLKEQPDN ALLMLYLGR LAYGRKLWKA
351 KGYLEASIAL KPSISARLVL AKVFDEIGEP QKAEARQNLV LEAVSDDERH
401 AALEQHS*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

40 ORF100 shows 93.5% identity over a 386aa overlap with an ORF (ORF100a) from strain A of *N. meningitidis*:

```

              10      20      30      40      50      60
orf100.pep  MKTVVWIVVLFAAAVGLALASGIYTGDOYIVLQOTMLRLNHAFLVGLSLIAVVWYFLFK
              |||
45 orf100a   MKTVVWIVVLFAAAVGLALASGIXTGDOYIVLQOTMLRLNHAFLVGLSLIAVVWYFLFK
              10      20      30      40      50      60
              70      80      90      100     110     120
50 orf100.pep FIIGVLNIPEKMRFGSARKGXKXKXLLALNKAAGLAYFEGRFKAELEASRVLVNKGVRDNR
              |||
orf100a      FIIGVLNXPPEKMRFGSARKGRKAAALNKAAGLAYFEGRFKAELEASRVLVNKGVRDNR
              70      80      90      100     110     120
              130     140     150     160     170     180
55 orf100.pep TLALMLXHAAGQMENIXRDRYLAELAKLPEKQQLSRVLLLAESALNRRDYEAEEANLH
              |||
orf100a      TLALMLGAHAAGQMENIELRDRYLAELAKLPEKQQLSRVLLLAESALNRRDYEAEEANLH
              130     140     150     160     170     180
60              190     200     210     220     230     240

```

	190	200	210	220	230	240
orf100.pep	AAAKMNNALTRLVRLKIRYAFDRGDALQVLAKTEPKLSKAGALGKSEMERYQNWAYRRQLA					
orf100a	AAAKMNNALTRLVRLQLRYAFDRGDALQVLAKTEKKSKAGAXGKSEMERYQNWAYRRQLA					
orf100.pep	DAADAALAKTKCLKRIKPSDLKNGELSVSVAEKYERLGLYADAVKWKVQHYPXNRPELLEA					
orf100a	DAADAALAKTKCLKRIKPSDLKNGELSVSVAEKYERLGLYADAVKWKVQHYPHNRPELLEA					
orf100.pep	FVESVRLFLGEREQQKAIDFDADWLKEQPNALLLMYLGLRLAYGRKLWGKAKGYLEASIAL					
orf100a	FVESVRLFLGERDQQKAIDFDADWLKEQPNALLLXYLGLRLAYGRKLWGKAKGYLEASIAL					
orf100.pep	KPSISARLVLTIKVFDEIGEPQKAEAH					
orf100a	KPSTISARLVLAKFVDETGEPOKAEAQNRNLVLSVAEENRPSAETHX					

The complete length ORF100a nucleotide sequence <SEQ ID 753> is:

1	ATGAAACGGG	TAGTCCTGGAT	ATGCTGCCTGT	TTGTCGGCGG	CNNTCGGGCT
51	GCATTTCGGG	TCCGGCATCTG	ACACGCGGGA	CTGTGTATATC	CTAGTCGGAC
101	AGACATCTCG	CAGAATCAAC	CTGCACACCT	TTGTGTTAAGT	TTGCGTACTAT
151	CGCGTCGTGG	TGTGGTATAT	CTGTCTCAAA	TTCATCATCG	TGGTACTCAA
201	TANCCCCGAA	AAGATGCAGC	TTTTCGGTGT	CGGCTGTAAA	GCGCGCAAA
251	CCTCGCTTGC	TTTGAACAG	CGGGTTTGG	CGATTTTGA	AGGGCGTTT
301	GAAATAGCGG	AACTTTGAAG	CTCCGCGGTA	TGTGGGAACAA	AAGAGCGCGG
351	GAGAACCGG	ACTTTTGACAT	TGATGTTGTC	CGCATCATGCC	CGCGGCGAGA
401	TCCGAATAACT	CGAGCTCGGC	GACGCTTATC	TTGCGGAATAT	CGCCAAACTG
451	TCGGAATAAGC	AGCAGCTTGC	CGTTTATAT	TTGTTTGGCG	AATCGGCGTT
501	GACRCGCGGC	GATATCAAG	CGCGGAGAGC	CAATCTTCAT	CGGGCGGCGT
551	AGATGAATGC	CAACTCTACG	CGCCTCGTC	TGTCGAACAT	TCGTTAGCGT
601	TTCGACAGGC	CGCAGCGGTT	CAGTGTCTTG	GCAAAACGAT	AAAAAATTTG
651	CRAAGCGGCG	CGCTNMGCGA	TTATCGGAAT	GCAACGCGAT	CAAAATTTGG
701	CATACCGCGC	CGACGTCNGA	GATGCTCGCG	ATGCGCGGAT	TTTGAAGAAC
751	TGCTCGAAGC	GGATTTCCGA	CAGCCTCAAA	ATACGCGGAC	TGAGCGTATC
801	GGTTCGCGAA	AGATATCAAC	TTTGGGACAT	GTATCGGCGT	CGGCTGATAT
851	GGCTCAGACA	CGATTATCG	CAACAACGTC	CGCGCGAACT	TTTGGAGCNC
901	TTTGTGCGAA	TTGCTGCTAT	TTTGGCGGAA	CGCGATCACT	AGAAAGAGCT
951	CGATTTGCTG	CGTCTTGGC	TGAAGAACGA	CGCGGATAT	CGGCTTCTGC
1001	TGANGTATCT	GCGTCGCGCT	GCTACGCGC	GCAAACTTTG	GCGCAAGCGA
1051	AAARGCTTAC	TTGAAGAGGAT	CATTGCATTA	ACGAGCAGTA	TTTCCGCGCG
1101	TTTGTGTTCT	CGAANGGTT	TTCGCGAACC	CGAGAGACGT	CATTAGACGCG
1151	AGCGACGCGC	CRACTTTGTT	TTGGCAGCGC	TTGCGCGAGA	AAACCGNCCT
1201	TCGCGCGGAA	CCGATTGGA			

This encodes a protein having amino acid sequence <SEQ ID 754>:

1	MTKVTVVVL	FAAXXGLA	SGMTGDSVY	VLQGMMLRN	LHAFVLGSLR
51	AVVVVYFLF	FIITVXVLP	KQIRXGDAV	KRQAAALAN	AGLAFYFGR
101	EKQELASRV	LGNKEAGDNR	TYALMGAHA	AGCMNIELR	DRYLAETAKL
151	PEKQELSLV	LIALSALNR	DIAELANLH	AAAKNNMIL	RLVLQLRYA
201	FRDGAQLVL	AKTEKXSVAG	AKGKSEMYR	QWYAYRQLX	DADAAALAKT
251	CLKRIQDQL	NGELXSVAG	KYERLGLGV	AKVWVKVGH	HNRRPELFL
301	FVSVPLGEG	KFQCARIDEA	ALLKLYRRL	ALLKLYRRL	AKRLKELG
351	KYGLAEATL	RSICARIDEA	AKVDETFEP	AKVDETFEP	LSAVSENRPR
401	SATHT				

ORF100a and ORF100-1 show 95.1% identity in 406 aa overlap:

orf100a.pep MKTVVWIVLVFAAAXGLALASGIXTGDVIYVLGGTMLRLNHAFLVGLSLIAVVVWYFLPK
| | | | |
orf100-1 MKTVVWIVLVFAAAVGLALASGIYTGDDIYVLGGTMLRLNHAFLVGLSLIAVVVWYFLPK

		10	20	30	40	50	60
5	orf100a.pep	FIIGVLN	PEKMRFGS	ARKGRKA	ALNKA	GLAYFEGR	FEKAELEASRV
	orf100-1	FIIGVLN	PEKMRFGS	ARKGRKA	ALNKA	GLAYFEGR	FEKAELEASRV
10	orf100a.pep	TLA	MLGAHAAGQ	MENTEL	RDRYLAE	IAKLPEK	QQLSRYLL
	orf100-1	TLA	MLGAHAAGQ	MENTEL	RDRYLAE	IAKLPEK	QQLSRYLL
15	orf100a.pep	AAAKMN	ANLTRL	VLRLQ	LYAFDR	GDALQ	VLAKTEK
	orf100-1	AAAKMN	ANLTRL	VLRLQ	LYAFDR	GDALQ	VLAKTEK
20	orf100a.pep	DAADA	AALKTK	LKRI	PD	SLKNGEL	SVSVAEK
	orf100-1	DAADA	AALKTK	LKRI	PD	SLKNGEL	SVSVAEK
25	orf100a.pep	FVES	SVFLG	EREQQ	KAI	DFAD	AWLKEQ
	orf100-1	FVES	SVFLG	EREQQ	KAI	DFAD	AWLKEQ
30	orf100a.pep	KPSI	SARL	VLAK	VFDE	TGEP	QKAE
	orf100-1	KPSI	SARL	VLAK	VFDE	TGEP	QKAE

Homology with a predicted ORF from *N.gonorrhoeae*

40 ORF100 shows 93.3% identity over a 386 aa overlap with a predicted ORF (ORF100ng) from

N.gonorrhoeae:

	orf100.pep	MKTVMVIIVLFAAAVGLALASGIYTGDDYIVLGGQTMLRLNLHAFVLGSLTAVVVWYFLFK	60
45	orf100ng	MKTVMVIIVLFAAAVGLALASGIYTGDDYIVLGGQTMLRLNLHAFVLGSLTAVVVWYFLFK	60
	orf100.pep	FIIGVLNPEKMRQGRFSARKGKKXLLALNKAGLAYFEGRFEKAELEASRVLVNKVGDRNR	120
	orf100ng	FIIGVLNIPENMRRSGSARKGRKKAALLNKAGLAYFEGRFEKAELEASRVLGNKAGDNR	120
50	orf100.pep	TLALMLXAAHAAGQMENTIXRDRYLAEIAKLPEKQQLSRYLLLAESALNRRDYEAEEANLH	180
	orf100ng	TLALMLGAHAAGQMENTELRDRYLAEIAKLPEKQQLSRYLLLAESALNRRDYEAEEANLH	180
55	orf100.pep	AAAKMNANLTRLVLRLQQLYAFDRGDALQVLAKTEKLSKAGALGKSEMERYQNWAYRQRLA	240
	orf100ng	AAAKMNANLTRLVLRLQLRYAFDRGDALQVLAKTEKLSKAGALGKSEMERYQNWAYRQRMQ	240
	orf100.pep	DAADAAALKTKCLKRIPDLSKNGELSVSVAEKYERLGLYADAVKWKQHYFXNRRPELLEA	300
60	orf100ng	DAADAAALKTKCLKRIPDLSKNGELSVSVAEKYERLGLYADAVKWKQHYPHNRRPELLEA	300
	orf100.pep	FVESVFLGGEREQQKAI DFADAWLKEQPDNALLMYLGRLAFGRKLWGKAGKYLEASIAL	360
65	orf100ng	FVESVFLGGEREQQKAI DFADSWLKEQPDNALLMYLGRLAYGKRLWGKAGKYLEASIAL	360

orf100.pep KPSISARLVLTQVFDEIGEPQKAEAH 386
 ||||| |||||:||||| :|: |||||:
 orf100ng KPSIPARLVLAQVFDETAQSQKAEQRNLVLASVAGENRPSAETR 405

The complete length ORF100ng nucleotide sequence <SEQ ID 755> is:

5 1 ATGAAACCG TAGTCTGSA TGTGTCTCG TTTGCCGCG CCGTCGGACT
 51 GCGCGTGGCT TCGGCGCATTT ACACCGGCGA CGTGATATAT GACTCGGACT
 101 AGACCATGCT CAGAATCAAC CTGCACGCGT TTGTGTTAGG TTCGCTGATT
 151 GCGCTCGTGG TGTGGTATTT CTTGTTTAAA TTCATCATCG CGCTACTCAA
 201 TATCCCCGAA AATATCGCGG GTTCCGGTTC GCGCGGAAA GCGCGCAAGG
 10 251 CCGCGCTTGC CTTGAATAAG GCGGGTTTGG CGTATTTTGA AGGCGGTTTT
 301 GAAAAGCGCG AACTCGAAGC CTCTCGAGTG TTGGGCAACA AAGAGGCGCG
 351 AGACAACCGG ACTTTGGCAT TGATGCTGGG CCGCGACGCG CGAGGACAGA
 401 TGGAAATAT CGAGCTCGCG GACCGTTATC TTGCGGAAAT CGCCAACTGT
 15 451 CCGGAAAAAC AGCAGCTTTC CCGCTATCTT CTGCTCGCGG AATCGCGCTT
 501 AARCCGCGCG GATTACGAAG CCGCGGAAGC CAATCTTCAI GCGCGGGAAT
 551 AGATGAATGC CAACCTTACG CGCCTCGTGG CTCTGCAACT TCGTTACGCC
 601 TTGATCGGGG CGCTTGGGCA AATCGGAAT GGAACGGTAT CAATTTGGG
 651 CAAGCGCGGG CGCTTGGGCA AATCGGAAT GGAACGGTAT CAATTTGGG
 701 CAATCGCGCG CCGATGCGGG GATGCTCGCG ATGCGCGCGC TTTGAAACAC
 20 751 TCGCTGAAGC GGATTCGCGA CAGCGTCAAA AACGGGGAAT TGagcGTATC
 801 GGTTCGCGAA AAGTACGAAC GTTTGGGACT GTATGCGCAT GCGGTCAAT
 851 GGGTCAAAACA CATTATCCG CACAACCGCC GCGCGAGCTT TTTGGAAGCC
 901 TTTTTCGAAA GCGTGCGCTT TTTGGCGGAG CGGCAACAGC AGAAAGCCAT
 951 CGATTTTGCC GATTCTTGGC TGAAGAACA CCGCGATAAC GCGCTCTGTC
 25 1001 TGATGTATCT CCGCGCGCTC GCCTACGGCC GCAAACTTTG GGGTAAGGCA
 1051 AAAGGCTACC TTGAAGCGAG TATTGCACTG AAGCCGAGTA TCCGCGCGCG
 1101 TTTGGTGTGT GCAAAAGGTT TTGACGAAC CGACAGTCG CAAAAGCGCG
 1151 AAGCACAGCG CAACCTGGTT TTGGCAAGCG TTGCGGGGGA AAACCGCCCT
 1201 TCCGCCGAAA CCCGTTGA

30 This encodes a protein having amino acid sequence <SEQ ID 756>:

35 1 MKTVVWIVVL FAAAVGLALA SGIYTGDOVY VLQOTMLRN LHAFVLGSLI
 51 AVVVVYFLFK FTTCVNLNPE NMRRSGSARK GKKAALNKL AGLAYFEGRF
 101 EKAELEASRV LGNKEAGDNR TLALMLGAHA AGQMENIELR DRYLAEIAKL
 151 PEKQQLSRYL LLAESALNRR DYEAAEANLH AAAMNANLTL RIVRLQLRYA
 201 FDRGDAQVLA ARTEKLSKAG ALGKSEMERV QNWAYRRQMA DAADAALKT
 251 CLKRIPTDSLK NGELSVSVAE KYERLGLYAD AVKWVQHYP HNNRPELLEA
 301 FVESVRFLEGE REQKALIDFA DSWLKEQPDN ALLMYLGR LAYGRKLWGKA
 351 KGYLEASIAL KPSIPARLV LAKVDETAQS QKAEQRNLV LASVAGENRP
 401 SAETR*

40 ORF100ng and ORF100-1 show 95.3% identity in 402 aa overlap:

		10	20	30	40	50	60
orf100-1.pep		MKTVVWIVVLFAAAVGLALASGIYTGDOVYIVLGQTMRLN	LHAFVLGSLI	AVVVVYFLFK	FTTCVNLNPE	NMRRSGSARK	GKKAALNKL
orf100ng		MKTVVWIVVLFAAAVGLALASGIYTGDOVYIVLGQTMRLN	LHAFVLGSLI	AVVVVYFLFK	FTTCVNLNPE	NMRRSGSARK	GKKAALNKL
		10	20	30	40	50	60
		70	80	90	100	110	120
orf100-1.pep		FIIGVLNIPKMQRFSGSARKGKKAALNKL	AGLAYFEGRF	EKAELEASRV	LGNKEAGDNR	TLALMLGAHA	AGQMENIELR
orf100ng		FIIGVLNIPENMRRSGSARKGKKAALNKL	AGLAYFEGRF	EKAELEASRV	LGNKEAGDNR	TLALMLGAHA	AGQMENIELR
		70	80	90	100	110	120
		130	140	150	160	170	180
orf100-1.pep		TLALMLGAHAAGQMENIELR	DRYLAEIAKL	PEKQQLSRYL	LLAESALNRR	DYEAAEANLH	AAAMNANLTL
orf100ng		TLALMLGAHAAGQMENIELR	DRYLAEIAKL	PEKQQLSRYL	LLAESALNRR	DYEAAEANLH	AAAMNANLTL
		130	140	150	160	170	180
		190	200	210	220	230	240
orf100-1.pep		AAAMNANLTL	RIVRLQLRYA	FDRGDAQVLA	ARTEKLSKAG	ALGKSEMERV	QNWAYRRQMA
orf100ng		AAAMNANLTL	RIVRLQLRYA	FDRGDAQVLA	ARTEKLSKAG	ALGKSEMERV	QNWAYRRQMA
		190	200	210	220	230	240

		250	260	270	280	290	300
5	orf100-1.pep	DAADAALKTKLKRIPD	SLKNGELSVSVAEKYERL	GLYADAVKWVKQHYPHNRRPELLEA			
	orf100ng	DAADAALKTKLKRIPD	SLKNGELSVSVAEKYERL	GLYADAVKWVKQHYPHNRRPELLEA			
		250	260	270	280	290	300
10	orf100-1.pep	FVESVRLGEREQQKAIDFADAWLKEQPDNALLMYLGLRAYGRKLWGKAKGYLEASIAL					
	orf100ng	FVESVRLGEREQQKAIDFADSWLKEQPDNALLMYLGLRAYGRKLWGKAKGYLEASIAL					
		310	320	330	340	350	360
15	orf100-1.pep	KPSISARLVLAQVDEIGEPQKAEAQRLNLVLEAVSDDERHAALQHSX					
	orf100n	KPSIPARLVLAQVDEITQSQKAEAQRLNLVLAQVAGENRPSAETRX					
		370	380	390	400		

Based on this analysis, including the presence of a putative leader sequence, a putative transmembrane domain, and a RGD motif, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 90

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 757>

```

1  ATGATGTTTT  CTGGTTCAA  GCTGTTTAC  TTGTTTTTG  TCATTTCTGTG
51  GTTTGCAGGG  CTGTTTTACC  TCCCGAGGAT  TTTCGTCAAT  ATGGCGATGA
101  TTGATGTGCC  CGCGGGCAAT  CCCGAGTATG  TCGCTCTGTC  GGGCATGGCG
151  GTGCGGCTGT  ACCGTTTTAT  GTCGCCGTTG  GGCTTCGGCG  CGGTCGTGTT
201  CGGCGCGGCG  ATACCGTTTG  CGCGCGGCTG  GTGGGCGAGC  GGCTGGGTAC
251  ACGTCAAAC  GTGTTTGGGC  TTGATGCTCT  TGGCTTACCA  GTTGATTGCG
301  GCGGTGCTGC  TGCGCCGTTT  TCAGGATTAC  AGCAATGCTT  TTTCACACCG
351  CTGGTACCGC  GTGTTCAACG  AAATCCCGTG  GCTGCTGATG  GTTGCCGCGC
401  TGTATstGGT  CGTGTTCAAA  CCGTTTTGA

```

This corresponds to the amino acid sequence <SEQ ID 758; ORF102>:

```

1  MMFSWFKLFH  LFFVISWFAG  LFYLERIFVN  MAMIDVPRGN  PEYVRLSGMA
51  VRLYREMSPL  GFGAVVFGAA  IFPAAGWWS  GWVHVKLCIG  LMLLAYQLYC
101  GVLLRRFQDY  SNAFSHRWYR  VNEIEVLLM  VAALYLVVEK  PF*

```

Further work revealed the complete nucleotide sequence <SEQ ID 759>:

```

1  ATGATGTTTT  CTGGTTCAA  GCTGTTTAC  TTGTTTTTG  TCATTTCTGTG
51  GTTTGCAGGG  CTGTTTTACC  TCCCGAGGAT  TTTCGTCAAT  ATGGCGATGA
101  TTGATGTGCC  CGCGGGCAAT  CCCGAGTATG  TCGCTCTGTC  GGGCATGGCG
151  GTGCGGCTGT  ACCGTTTTAT  GTCGCCGTTG  GGCTTCGGCG  CGGTCGTGTT
201  CGGCGCGGCG  ATACCGTTTG  CGCGCGGCTG  GTGGGCGAGC  GGCTGGGTAC
251  ACGTCAAAC  GTGTTTGGGC  TTGATGCTCT  TGGCTTACCA  GTTGATTGCG
301  GCGGTGCTGC  TGCGCCGTTT  TCAGGATTAC  AGCAATGCTT  TTTCACACCG
351  CTGGTACCGC  GTGTTCAACG  AAATCCCGTG  GCTGCTGATG  GTTGCCGCGC
401  TGTATCTGGT  CGTGTTCAAA  CCGTTTTGA

```

This corresponds to the amino acid sequence <SEQ ID 760; ORF102-1>:

```

1  MMFSWFKLFH  LFFVISWFAG  LFYLERIFVN  MAMIDVPRGN  PEYVRLSGMA
51  VRLYREMSPL  GFGAVVFGAA  IFPAAGWWS  GWVHVKLCIG  LMLLAYQLYC
101  GVLLRRFQDY  SNAFSHRWYR  VNEIEVLLM  VAALYLVVEK  PF*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with HP1484 hypothetical integral membrane protein of *H. pylori* (accession number AE000647)

ORF102 and HP1484 show 33% aa identity in 143aa overlap:

```

5      orf102  3  FSWFKL FHLFFVISW FAGLFYLPRI FVNMA MIDVPRGNPEYVRLSGMAVRLYR FMSPLGF 62
      HP1484  8  FLWVKAFHVI AVISWMAALFYLPRLFVYHAENAHKKEPFGV VQIQEK--KLYSFIAS PAM 65

      orf102  63  GAVVFGAAIPFAAG---WWSGSGVHVHVKLCGLGMLLAYQLYCGVLLRRFPQDYSNAFSHRWY 119
      HP1484  66  G + + + GW+H KL L ++LLAY YC +R + + R+Y
      110      66  GFTLTITGILMLLIEPTLFKSGGWLHAKLALVLLLAYHFYCKKCKRELEKDPTRNRARFY 125

      orf102  120 RVFNEIPXXXXXXXKXFKPF 142
      HP1484  126 RVFNE P KPF
      15      126 RVFNEAPTILMILIVILVVVKPF 148
  
```

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF102 shows 99.3% identity over a 142aa overlap with an ORF (ORF102a) from strain A of *N.*

meningitidis:

```

20      orf102.pep  MMFSWFKLFHLFFVISW FAGLFYLPRI FVNMA MIDVPRGNPEYVRLSGMAVRLYR FMSPL
      orf102a      MMFSWFKLFHLFFVISW FAGLFYLPRI FVNMA MIDVPRGNPEYVRLSGMAVRLYR FMSPL
      25      10      20      30      40      50      60
      orf102.pep  CFCGAVVFGAAIPFAAGWWSGSGVHVHVKLCGLGMLLAYQLYCGVLLRRFPQDYSNAFSHRWYR
      orf102a      CFCGAVVFGAAIPFAAGWWSGSGVHVHVKLCGLGMLLAYQLYCGVLLRRFPQDYSNAFSHRWYR
      30      70      80      90      100      110      120
      orf102.pep  VFNEIPVLLMVAALYXVVFVKPF
      orf102a      VFNEIPVLLMVAALYXVVFVKPF
      35      130      140
  
```

The complete length ORF102a nucleotide sequence <SEQ ID 761> is:

```

1      ATGATCTTTT  CITGGTTCAA  GCTGTTTCAC  TTGTTTTTTG  TCATTTTCGTG
51     GTTTCAGCG  CTGTTTTAC  TGCCGAGSAT  TTTCGTCAAT  ATGGCGATGAA
101    TTGATCTGCC  CGCGGCAAT  CCGAGTATG  TGCGTCTGTC  GGCGATGGCG
40     GTGCGGCTGT  ACCGTTTAT  GTGCGCGTGT  GCGTTCGCGC  GCGTTCGCTT
151    CGGCGCGCGC  ATACCGTTTG  CGCGCGCGTG  GTGGGCGCAG  GCGTGGGTAC
201    ACGTCAAACT  GTGTTTGGGC  TTGATGCTCT  TGGCTTACCA  GTTGATTATG
251    GCGTGTCTGC  TGCGCGGTTT  TCAGGATTAC  AGCAATGCTT  TTTACACCG
301    CTGTACCGC  GTGTTCAACG  AAATCCCGT  GCTGCTGATG  GTTGC CGCGC
351    TGTATCTGGT  CGTGTTCAAA  CCGTTTGA
45     401
  
```

This encodes a protein having amino acid sequence <SEQ ID 762>:

```

1      MMFSWFKLFH  LFFVISW FAG  LFYLPRI FVN  MAMIDVPRGN  PEYVRLSGMA
51     VRLYR FMSPL  GFCGAVVFGAA  IPFAAGWWSG  GWHVHVKLCGL  LMLLAYQLYC
101    GVLRRFPQDY  SNAFSHRWYR  VFNEIPVLLM  VAALYL VVVF  PF*
  
```

ORF102a and ORF102-1 show complete identity in 142 aa overlap:

```

50      orf102a.pep  MMFSWFKLFHLFFVISW FAGLFYLPRI FVNMA MIDVPRGNPEYVRLSGMAVRLYR FMSPL
      orf102-1      MMFSWFKLFHLFFVISW FAGLFYLPRI FVNMA MIDVPRGNPEYVRLSGMAVRLYR FMSPL
      55      10      20      30      40      50      60
  
```

```

              70      80      90      100      110      120
orf102a.pep  GFGAVVFGAAIPFAAGWWSGWSVHVHVKLCGLMMLLAYQLYCGVLLRRFQDYSNAFSHRWYR
5 orf102-1    GFGAVVFGAAIPFAAGWWSGWSVHVHVKLCGLMMLLAYQLYCGVLLRRFQDYSNAFSHRWYR
              70      80      90      100      110      120

              130      140
orf102a.pep  VFNEIPVLLMVAALYLNVFKPF
10 orf102-1   VFNEIPVLLMVAALYLNVFKPF
              130      140

```

Homology with a predicted ORF from *N.gonorrhoeae*

15 ORF102 shows 97.9% identity over a 142 aa overlap with a predicted ORF (ORF102ng) from *N. gonorrhoeae*:

```

orf102.pep  MMFSWFKLHLEFFVISWIFAGLFYLPRI FVNMMAMIDVPRGNPEYVRLSGMAVRLYRFMSPL 60
orf102ng    MMFSWFKLHLEFFVISWIFAGLFYLPRI FVNMMAMIDAPRGNPEYVRLSGMAVRLYRFMSPL 60

orf102.pep  GFGAVVFGAAIPFAAGWWSGWSVHVHVKLCGLMMLLAYQLYCGVLLRRFQDYSNAFSHRWYR 120
orf102ng    GFGAVVFGAAIPFAAGWWSGWSVHVHVKLCGLMMLLAYQLYCGVLLRRFQDYSNAFSHRWYR 120

orf102.pep  VFNEIPVLLMVAALYLNVFKPF 142
orf102ng    VFNEIPVLLMVAALYLNVFKPF 142

```

The complete length ORF102ng nucleotide sequence <SEQ ID 763> is:

```

1  ATGATGTTTT CTGCGTTCAA GCTGTTTCAC TTGTTTTTG TCATTTCGTG
30 51  GTTTCGACGG CTGTTTTACC TGCCGAGGAT TTGCGTCAAT ATGGCGATGA
101 TTGATGCGCC GCGCGGCAAT CCGAGTATG TGCGCCTGTC GGGGATGCGC
151 GTGCGGTGTG ACCGTTTTAT GTCCGCTTTG GGTTCGCGC CGGTGCTGTT
201 CGGCGCGGCG ATACCGTTTG CCGCGggcgg GTGGGCGcgc ggctggGTTT
251 ACGTCAAAC TGTTTTGGGC TTGATGCTCT TGGCTATCA GTTGATTGCG
301 GCGCTGCTGC TGCGCCGTTT TCAGGATTAC AGCAATGCTT TTTCACACCG
35 351 CTGGTACCGC GTGTTCAaag aAATCCCGT GCTGCTGATG GTTGCCGCGC
401 TGTATCTGTT CGTGTTCAAA CCGTTTTGA

```

This encodes a protein having amino acid sequence <SEQ ID 764>:

```

1  MMFSWFKLFH LFFVISWIFAG LFYLPRI FVN MAMIDAPRGN PEYVRLSGMA
40 51  VRLYRFMSPL GFGAVVFGAA IPFAAGWWS GWVHVHVKLCGL LMLLAYQLY
101 GVLLRRFQDY SNAFSHRWYR VFNEIPVLLM VAALYLNVFK PF*

```

ORF102ng and ORF102-1 show 98.6% identity in 142 aa overlap:

```

              10      20      30      40      50      60
orf102-1.pep MMFSWFKLHLEFFVISWIFAGLFYLPRI FVNMMAMIDVPRGNPEYVRLSGMAVRLYRFMSPL
45 orf102ng    MMFSWFKLHLEFFVISWIFAGLFYLPRI FVNMMAMIDAPRGNPEYVRLSGMAVRLYRFMSPL
              10      20      30      40      50      60

              70      80      90      100      110      120
orf102-1.pep GFGAVVFGAAIPFAAGWWSGWSVHVHVKLCGLMMLLAYQLYCGVLLRRFQDYSNAFSHRWYR
50 orf102ng    GFGAVVFGAAIPFAAGWWSGWSVHVHVKLCGLMMLLAYQLYCGVLLRRFQDYSNAFSHRWYR
              70      80      90      100      110      120

              130      140
orf102-1.pep VFNEIPVLLMVAALYLNVFKPF
55 orf102ng    VFNEIPVLLMVAALYLNVFKPF
              130      140

```


In addition, ORF102ng shows significant homology to a membrane protein from *H. pylori*:

```
gi|2314656 (AE000647) conserved hypothetical integral membrane protein
[Helicobacter pylori] Length = 148
Score = 79.2 bits (192), Expect = 1e-14
Identities = 50/147 (34%), Positives = 68/147 (46%), Gaps = 13/147 (8%)

Query: 3   FSWFKLHLFFVISWFAGLFYLERIFVNMAMIDAFRGNPEYVRLSGMAVRLRYFMSPLGF 62
          F W K FH+ VISW A LFYLER+ FV A + V++ +LY F++
Sbjct: 8   FLWVKAFHVIAVISWMAALFYLERLFVYHAENAHKKEFVGQVIQEK--KLYSTIASPAM 65

Query: 63   GAVVFGAARIP-----FAAGRWGSGWVHVKLCLGLMLLAYQLYCGVLLRRFDQSNAPS 115
          G + + F +G GN+H KL L ++LLAY YC +R + +
Sbjct: 66   GFTLLTGILMLLIEPTLFKSG----GNLHAKIALVLLLAYHFYCKCKMRELEKDPTRRN 121

Query: 116  HRWYRVFNEIPXXXXXXXXXXXXXFKPF 142
          R+YRVFNE P KPF
Sbjct: 122  ARFYRVFNEAPTILMILIVILVVVKPF 148
```

Based on this analysis, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 91

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 765>:

```
1   ATGGCAAAAA TGATGAAATG GCGGGCTGTT GCGGCGTCG CCGCGCACG
51  GGTTTGGGGC GGATGCTCTT AACTGAAGCC CGAGCGCGAC GTGCTTGATA
101 TTACGGAAAC GGTACGGCGC GGC // .....
25  //.. ATTTCTGTTA CGATTTTGTC CGAACCGGAT ACGCCGATTA AGCGGAAGCT
51  CGACAGCGTC GACCCCGGGC TGACCAAGAT GTGCTGGGGC GGTTCACACA
101 GCAGTACGGA TACGGCTTCC AATGCGGCTC ACTATTATGC CGTTCGTTT
151 GTGCCGAATC CGGACGGCAA ACTCGCCACG GGGATGACGA CGCAGAATAC
30  GGTGGAATTC GACGGCGTGA AAAATGTGCT GATTATTCCG TCGCTGACCG
251 TGAAAAATCG CCGCGGCAAG GCGTTTGTGC GCGTGTGGG TCGGAGCGGC
301 AAGCGCGGGC AACCGGAAAT CCGGACCGGT ATGAGAGACA GTATGAATAC
351 CGAAGTAAAA AGCGGGTTGA AAGAGGGGGA CAAAGTGTCT ATCTCCGAAA
401 TAACCGCCGC CGAGCACAAG GAAAGGGCGC AACGCGCCT AGCGCGCCGC
451 CCGCGCGGAT AA
```

This corresponds to the amino acid sequence <SEQ ID 766; ORF85>:

```
1   MAKMKKAAAV AAVAAAAVMG GWS.LKPEPH VLDITETVRR G.....
51  .....
101  .....
151  .....
40  201  .....I SFTILSEPT
251  PIKAKLDSVD PGLTMMSSGG YNSSTDASN AVYYIARSFV PNPDGKLATG
301  MTTQNTVEID GVRNVLIPS LTVKNRGGKA FVRVLGADGK AAEREIRTCM
351  RDSMNTVEKS GLKEGDKVVI SEITAEQOE SGERALGGPP RR*
```

Further work revealed the further partial nucleotide sequence <SEQ ID 767>:

```
45  1   ..GTATCGGTG GCGCGCAGGC ATCGGGGCGC ATTAAGATAC TTTATGTCAA
51  ACTCGGGCAA CAGGTAAAAA AGGGCGATTT GATTGCGGAA ATCAATTCGA
101  CCTCGCAGAC CAATACGCTC AATACGGAAA AATCAAGTTT AGAAACGTAT
151  CAGCGCAGAC TGTGTCTGCG ACAGATTGCA TTGGCGACGC CGGAGAAGAA
50  201  ATATAGACGT ACGCGGCGCT TATCGAGAGC CCGCAGAGC CAATGTCC
251  ATTTGGAAGG CGCGCAGAT GCGTTTGCCG CGCGCAGAGC CAATGTCC
301  GAGCTGAAGG CTTTAAATCAG ACAGAGCAAA ATTTCCATCA ATACCGGGA
351  GTCGGAATTG GGCTACACGC GATTAACCGC AACGATGGAC GCACCGGTGG
401  TGGCGATTCT CGTGGAAAGG GGGCAGACTG TGAACCGCGC CGACTCTACG
451  CCGACGATTG TCCAATTGGC GAATCTGGAT ATGATGTTGA ACAAAATGCA
55  501  GATTGCCGAG GCGCATATTA CCAAGGTGAA GCGCGGGCAG GATATTTCTG
551  TTACGATTTT GTCGGAACGC GATACGCCGA TTAAGCGGAA GTCGCAACGC
601  GTCGACCCCG GGCTGACCAC GATGTGCTCG GCGCGTTACA ACAGCAGTAC
```

651 GGATACGCGCT TCCATTCGGG TCCTCATTTA TGCCCGCTTCG TTTGTGGCGA
701 ATCCGCGAGCG CAACTCGCGG CAGCGCAGATGA CACGCGAGAAA TACGGTTGCA
751 ATCCGAGCGCG TGCAAAATGCT CGGTATTTT CCGTGCGTGA CCGTGAAGAAA
801 CCGGACCGCGC AAGCGCTTTG TCGCGGCTTT GCGTGCGCAG CGCGAAGCGCG
851 CGGACGAGCGA AATCCGCGACC GGATGAGAGC ACAGTATGATG CCGTGAAGTAT
901 AAAAGCGCGGT TGAAGAAGGG GCGTAAGAATG GTCATCTCCG AAATAACCGCG
951 CCGCGAGCAA CAGGAAGCG GCGCAAGCGC CCTAGGCGCG CCGCGCGCGC
1001 GATAA

This corresponds to the amino acid sequence <SEQ ID 768; ORF85-1>:

```

1      ..VSVGQAASQG IKILYVKLQG QVKKGDILAE INSTSQTNWL NTEKSKLEYT
51     QAKLVYISQIA LGSAAEKYKR QAAWKENATD SKEDLESQD AFAAKANVA
101    ELKALIRIQA SIINTAESL GYTRITATMD GTVUAILVE QGTVNAAQS
151    PTVOLANLD MMLNKMQIAE GDITVYKAG DIFSTLFE DPTTKARLDS
201    VFDGPTTMSG GYNSNSTDA SNNAVYKAS FVNPNDGKLA TGPMIQTQV
251    IDGVKNVLNI PSLTVKNRGG KAFVVLVGAD GKAAERERT GMRDSMMNTV
301    KSLGREGDKV VISEITAAEG QSEGRALGR PRR*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF85 shows 87.8% identity over a 41aa overlap and 99.3% identity over a 153aa overlap with

an ORF (ORF85a) from strain A of *N. meningitidis*:

```

                                10      20      30      40
orf85.pep    MAKMMKWAAVAATAAAVAVGGWS-LKEEPEHLDITETVRRG
              |||||
orf85a       MAKMMKWAAVAATAAAVAVGGWSLKPEEQAAYITETVRRGDISRTVSATGEISPSNLVS
              10      20      30      40      50      60
                                //
                                80      90      100
orf85.pep    .....TSFTLSEIPDTPIKAKLDSVDPLGTTMSSG
orf85a       TIVQLANLDMMLNKMQIAEGDITVKAGQDISFTLSEIPDTPIKAKLDSVDPLGTTMSSG
              210     220     230     240     250     260
                                110     120     130     140     150     160
orf85.pep    GYNSSTDASNAVYYARSFVNPDPGLKATGMTTQNTVEIDGVKNVLIIPLSTVKNRGGK
              |||||
orf85a       GYNSSTDASNAVYYARSFVNPDPGLKATGMTTQNTVEIDGVKNVLIIPLSTVKNRAGGR
              270     280     290     300     310     320
                                170     180     190     200     210     220
orf85.pep    AFVVRVLGADGKAAREIRTMGRDSMNTVEKVGSLKEGDKVVISITAAPQESGERALGGP
              |||||
orf85a       AFVVRVLGADGKAAREIRTMGRDSMNTVEKVGSLKEGDKVVISITAAPQESGERALGGP
              330     340     350     360     370     380
                                230
orf85.pep    PRRX
              |||
orf85a       PRRX
              390

```

The complete length ORF85a nucleotide sequence <SEQ ID 769> is:

1	ATGGCAAAAG	GGATGAATGT	GGCGCTGTGT	GGCGCGGGTG	CGGCGCGGAC
51	GGTTTGGGCG	GGATGGCTCT	ATTCAACGAC	CGAGCGCGAG	CGTCTGTATC
101	TTACGGAAAT	CGCTAGCGCG	GGCGACATCA	CGCGGACGGT	TTCTTCGACA
151	GGGAGAGATT	CGCGCTCCAA	CCTGGTATCG	GTTCGGCGGAG	AGGCATCGGG
201	CGRAGTAACT	AAACTTTTAGT	TCGAACCTCG	CGACACAGTT	AAAAAGGGCG
251	ATTGATTGTC	GGAATCAAT	TGCAGCTCGC	AGACCAATAC	CGTCAATGAT
301	AAAAATATCA	ATTGGAAGAC	GATATCGCGC	AGGCTGTGGT	CGGCAATGAT
351	TGCAATGGGC	AGCGCGAGAA	GGAAATATAA	GCGTTCAGCG	CGGCTTGAGA
401	AGGATGATG	GACCGCTAAA	GAGAAATTTG	AAAGCGGCGA	GAGTCGGCTT
451	CGCGCGCCCA	AAGCCAATGT	TGCGGAGCTG	AGAGCTCTAA	TCGACAGAG

501 CAAAATTTCC ATCAATACCG CCGAGTCGGA ATTGGGCTAC ACGCGCATT
 551 CCGCAACGAT GACCGGCACG GTGGTGGCGA TTCTGTGGGA AGAGGGGGAG
 601 ACTCTGACAG CCGCGCGATC TACCGCGACG ATTCTCCGAT TGGCGAACT
 651 GGSATATGATG TTGACAAAGA TGCAGATTGC CGAGGGCGAT ATTACCAAGG
 701 TGAAGGCGGG CGAGGATATT TCGTTTACGA TTTTGTCCGA ACCGGATACG
 751 CCGATTAAAG CGAAGCTCGA CAGCGTCGAC CCGGGGCTGA CCACGATGTC
 801 GTCCGGGCGG TACAACAGCA GTACGGATAC GGCTTCCAAT CGGCTCTACT
 851 ATTATGCCCG TTGCTTTTGT CCGAATCCGG ACGGCAAACT CGCCACGGGG
 901 ATGACGACGC AGAATACGGT TGAATTCGAC GGTGTGAAAA ATGTGCTGAT
 951 TATTCCGCTCG CTGACCGTGA AAAATCCGGC CGGCAGGGCG TTTGTGCCGG
 1001 TGTTTGGTGC AGACGGCAAG CGCGCGGAAC GCGAAATCCG GACCGCTATG
 1051 AGAGACAGTA TGAATACCGA AGTAAAAAGC GGGTTGAAG AGGGGGACAA
 1101 AGTGGTCATC TCGAAATAA CCGCCGCCGA CGACGAGGAA AGCGGCGAAC
 1151 GCGCCCTAGG CGGCCGCCG CGCCGATAA

15 This encodes a protein having amino acid sequence <SEQ ID 770>:

1 MAKMMKVAAR AAVAAAIVWG GWSYLPKPEQ AAYITTVRR GDISRTVSAT
 51 GEISPSNLVS VQAQASGGQIK KLYVKLGQQV KRGDLIAEIN STSQTNTINT
 101 EKSKLETYQA KLVSAQIALG SAERKVKRQA ALWKDDATAK BDLESQDAL
 151 AAKANVAEL KALIROSKI INTAESELG TRITATMDGT VVAILEEQQ
 201 TVNAAQSTFT IVQLANLDM LNMQIAEGD ITKVKAGQDI STTILSEPD
 251 PIKAKLDSVD PGLTMTSSGG YNSSTDASN AVYYIARFV PNPDKLATG
 301 MTTQNTVEID GVKNVLIIPS LTVKNRGGRA FVRVLGADGK AAEREIRTMG
 351 RDSMNTVEKS GLKEGDKVVI SEITAAEQQE SGERALGGPP RR*

ORF85a and ORF85-1 show 98.2% identity in 334 aa overlap:

25	orf85a.pep	30	40	50	60	70	80
		PQAAAYITETVRRGDISRTVSATGEISPSNLVS	VQAQASGQIKKLYVKLGQQVKKGDLIAE				
	orf85-1						
					VSVGAQASGQIKLYVKLGQQVKKGDLIAE		
30					10	20	30
	orf85a.pep	90	100	110	120	130	140
		INSTSQTNTINTKESKLETTQAKLVSAQIALGSAE	KKRQAALWKDDATAKBDLESQAQD				
	orf85-1						
		INSTSQTNTINTKESKLETTQAKLVSAQIALGSAE	KKRQAALWKENATSKEDLESQAQD				
35		40	50	60	70	80	90
	orf85a.pep	150	160	170	180	190	200
		ALAAAKANVAELKALIROSKIINTAESELGYTRITATMDGT	VVAILEEQQTVNAAQST				
	orf85-1						
		AFAAAKANVAELKALIROSKIINTAESELGYTRITATMDGT	VVAILEEQQTVNAAQST				
40		100	110	120	130	140	150
	orf85a.pep	210	220	230	240	250	260
		FTIVQLANLDMMLNMQIAEGDITKVKAGQDISFTTILSEPD	TPIKAKLDSVDPGLTMTSS				
	orf85-1						
		FTIVQLANLDMMLNMQIAEGDITKVKAGQDISFTTILSEPD	TPIKAKLDSVDPGLTMTSS				
45		160	170	180	190	200	210
	orf85a.pep	270	280	290	300	310	320
		GGYNSSTDASNNAVYYIARFVNPDPDKLATGMTTQNTVEID	GVKNVLIIPSLTVKNRGG				
	orf85-1						
		GGYNSSTDASNNAVYYIARFVNPDPDKLATGMTTQNTVEID	GVKNVLIIPSLTVKNRGG				
50		220	230	240	250	260	270
	orf85a.pep	330	340	350	360	370	380
		RAFVRVLGADGKAAEREIRTMGRDSMNTSVKSLKEGDKRV	ISEITAAEQQESGERALGG				
	orf85-1						
		KAFVRVLGADGKAAEREIRTMGRDSMNTSVKSLKEGDKRV	ISEITAAEQQESGERALGG				
55		280	290	300	310	320	330
	orf85a.pep	390					
		PPRRX					
	orf85-1						
60							

65 Figure 19D shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF85a..

Homology with a predicted ORF from *N.gonorrhoeae*

ORF85 shows a high degree of identity with a predicted ORF (ORF85ng) from *N.gonorrhoeae*:

5	ORF85	1	MAKMMKWAAVA AAAAAVWGGWS.LKPEPHVLDTITVRRG.....	40
	ORF85ng	1	MAKMMKWAAVA AAAAAVWGGWSYLKPEPQAAAYTEAVRRGDISRTVSAT	50
10	ORF85	ISFTILSEPD	250
	ORF85ng	201	TVNAAQSTPTIVQLANLDMMLNMQIAEGDITKVKAGQDISFTILSEPD	250
15	ORF85	251	PIKAKLDSVDPLTTMSSGGYNSSTDASNAVYYARSFVFPNPDGKLATG	300
	ORF85ng	251	PIKAKLDSVDPLTTMSSGGYNSSTDASNAVYYARSFVFPNPDGKLATG	300
20	ORF85	301	MTTQNTVEIDGVKNVLLI PSLTVKNRGGKAFVRVLGADGKAEREIRTM	350
	ORF85ng	301	MTTQNTVEIDGVKNVLLI PSLTVKNRGGKAFVRVLGADGKAEREIRTM	350
25	ORF85	152	RDSMNTVEKSGLKEGDKVISEITAAEQQESGERALGGPFRR	393
	ORF85ng	351	KDSMNTVEKSGLKEGDKVISEITAAEQQESGERALGGPFRR	393

The complete length ORF85ng nucleotide sequence <SEQ ID 771> is:

25	1	ATGGCAAAAA	TGATGAAATG	GGCGGCTGTT	GGCGCGCTGC	CGCGCGCaac
	51	GSTTTGGGGC	GGATGGTCTT	ATCTGAAGCC	CGAACCGCAG	GCTGCTTATA
30	101	TTACGGAaac	ggTCAGCGCG	GGCGATATCA	GCGCGAGCGT	TTCGCGAGCG
	151	GtcgagATT	CGCGTCCAA	CCTGGATGC	GTCCGCGCGC	AGGCTTCGGG
35	201	CGCAGATTAA	AAGCTTTATG	TCAACTCGG	GCACACGCTG	AAAAAGCGCG
	251	ATTTGATTGC	GGAAATCAAT	TCGACCAAGC	AGACCAACAC	GATCATATAT
40	301	GAAAAATCCA	AATTGGAAC	GTATCAGCGC	AAGCTGGTGT	CGGCACAGAT
	351	TGCATTGGCG	AGCGCGGAGA	AGAAATATAA	CGCTCAGCGC	CGGTTGTGGA
45	401	AGGATGATGC	GACCTCTAAA	GAAAGTTTG	AAAGCGCGCA	GGAATGCGCTT
	451	CGCGCGCGCA	AAGCCAATGT	TGCCAGATTG	AAGGCTTTAA	TCAGACGAGAG
50	501	CAAAATTTCC	ATCAATACCG	CGAGTCGGA	TTTGGGCTAC	ACGCGCATTA
	551	CCGCGACGAT	GGACGCGACG	GTGGTGGGCA	TTCCTGTGGA	AGAGGGGCGAG
55	601	ACTGTGAACG	CGGCGCAGTC	TACGCGCAGC	ATTGTCCAAT	TGGCGAATCT
	651	GGATATGATG	TTGAACAAAA	TGCAGATTGC	CGAGGGCGAT	ATTACCAAGG
60	701	TGAAGCGGGG	GCAGGATATT	TCGTTTACGA	TTTGTTCGGA	ACCGGATACG
	751	CGATTAAGG	CGAGCTCGCA	CAGCGTCGAC	CCCGGGCTGA	CCACGATGTC
65	801	GTCGGCGGCG	TACACACGA	CGGCGATGAT	GGCTTCACAT	CCGCTCTAGT
	851	ATTATGCGCG	TTCTTTGTG	CCGAATCCGG	ACGCGCAACT	CGCCACGCGG
70	901	ATGACGACGC	AGAATACGTT	TGAATCCGAC	GGTCTCAAAA	ATGTGTTGCT
	951	TATTCCGTCG	CTGACCGTGA	AAAAATCGCG	CGCGAAGCGC	TTCTGACGCG
75	1001	TGTTGGGTGC	GGACGGCAAG	GCAGTGAAC	CGGAAATCCG	GACCCGATATG
	1051	AAAGACAGTA	TGAATACCGA	AGTGAAGAAG	GGGTTGAAGG	AGGGGGGAGT
80	1101	AGTGTGTCAT	TCCGAATATA	CCGCGCGCGA	GCAGACAGAA	AGCGCGCAAG
	1151	GCGCCCTAGG	CGGCGCGCGG	CGCCGATAA		

This encodes a protein having amino acid sequence <SEQ ID 772>:

50	1	MAKMMKWAAY	AAAAAAVWG	GWSYLKPEPO	AAYITEAVRR	GDISRTVSAT
	51	GEISPSNLVS	VGAQASGQIK	KLYVKLGQV	KKGLDLIAE	INTQTNTIDM
55	101	EKSLETLYQA	KLVSQIALG	SAEKYKRQA	ALWKDDATSK	EDLESQAQDL
	151	AAKANVAEL	KALIRSGRTS	INTASDLYT	TRTATMDGT	VVAIPVEEGQ
60	201	TVNAAQSTPT	IVQLANLDM	NKMQIAEGD	ITKVKAGQDI	SFTILSEEDT
	251	PIKAKLDSVD	PGLTTMSSGG	YNSSTDASN	AVYYARSFV	ENPDGKLATG
65	301	MTTQNTVEID	GVKNVLLIPS	LTVKNRGGKA	FVRVLGADGK	AVEREIRTM
	351	KDSMNTVEKS	GLKEGDKVVI	SEITAAEQQE	SGERALGGP	RR*

ORF85ng and ORF85-1 show 96.1% identity in 334 aa overlap:

60	orf85ng	30	40	50	60	70	80
	orf85-1	PQAAAYITETVRRGDISRTVSATGEISPSNLVS	VGAQASGQIKKLVKLGQVKKGGLIAE				
				VSVGAQASGQIKILYVKLGQVKKGGLIAE			

					10	20	30
5	orf85ng	90	100	110	120	130	140
		INSTTQTNTIDMEKSKLETYQAKLVSAQIALGSAEKKYKQAAALWKDDATSKEDLESQAQD					
	orf85-1	40	50	60	70	80	90
		INSTSQNTNTLITEKSKLETYQAKLVSAQIALGSAEKKYKQAAALWKENATSKEDLESQAQD					
10	orf85ng	150	160	170	180	190	200
		ALAAAKANVAELKALIRQSKISINTAESDLGYTRITATMDGTVVVAIPVEEGQTVNAAQST					
	orf85-1	100	110	120	130	140	150
		AFAAAKANVAELKALIRQSKISINTAESDLGYTRITATMDGTVVVAIPVEEGQTVNAAQST					
15	orf85ng	210	220	230	240	250	260
		PTIVQLANLDMMLNKMQIAEGDITKVKAGQDISFTILSEPDTPIKAKLSDVDPGLTTMSS					
	orf85-1	160	170	180	190	200	210
		PTIVQLANLDMMLNKMQIAEGDITKVKAGQDISFTILSEPDTPIKAKLSDVDPGLTTMSS					
20	orf85ng	270	280	290	300	310	320
		GGYNSSTDTSNAVYYARSFVNPDPGKLGATGMTTQNTVEIDGVKNVLLIPSLTVKNRGG					
	orf85-1	220	230	240	250	260	270
		GGYNSSTDTSNAVYYARSFVNPDPGKLGATGMTTQNTVEIDGVKNVLLIPSLTVKNRGG					
25	orf85ng	330	340	350	360	370	380
		KAFVRVLGADGKAVEREIRTGMKDSMNTVEKSGLKEGDKVVISSEITAAEQQESGERALGG					
30	orf85-1	280	290	300	310	320	330
		KAFVRVLGADGKAAREIRTGMKDSMNTVEKSGLKEGDKVVISSEITAAEQQESGERALGG					
35	orf85ng	390					
		PPRRX					
	orf85-1	390					
		PPRRX					

In addition, ORF85ng shows significant homology to an *E. coli* membrane fusion protein:

40	gi 1787104 (AE000189) o380; 27% identical (27 gaps) to 332 residues from membrane fusion protein precursor, MTRC_NEIGO SW: P43505 (412 aa) [Escherichia coli] Length = 380 Score = 193 bits (485), Expect = 2e-48 Identities = 120/345 (34%), Positives = 182/345 (51%), Gaps = 13/345 (3%)
45	Query: 29 PQAAIYITVRRGDISRTVSATGEISPSNLVSGAQSAGQIKKLYVKLGQVKKGGDLIAE 88 P Y T VR GD+ ++V ATG++ V VQAQ SQG+K L V +G +VKK L+ Sbjct: 41 PVPTYQTFLVIRPGDLQSSVLATGKLDALRKVDVGAQVSGQLKTLVSAIGDKVKKQDLGV 100
50	Query: 89 INSTTQTNTIDMEKSKLETYQAKLVSAQIALGSAEKKYKQAAALWKDDATSKEXXXXXX 148 I+ N I ++ L +R+ A+ L A Y RQ L + A S++ Sbjct: 101 IDPEQAENQIKEVEATIMELRAQRQQAELKLARVITYSRQQLRAQTKAVSQDLDATA 160
55	Query: 149 XXXXXXXXXXXXXXXXIRQSKISINTAESDLGYTRITATMDGTVVVAIPVEEGQTVNAAQST 208 I++++ S++TA+++L YTRI A M G V I +GOTV ARQ Sbjct: 161 EMAVKQAQIGTIDAQIKRNOASLDATAKTNLDYTRIVAPMAGEVTQITTLQQQTVIAAQQ 220
60	Query: 209 PTIVQLANLDMMLNKMQIAEGDITKVKAGQDISFTILSEPDTPIKAKLSDVDPGLTTMSS 268 P I+ LA++ ML K Q++E D+ +K GQ FT+L +P T + + + V P Sbjct: 221 PNILTLADMSAMLVKAQVSEADVHLKPGQKAWFTVLGDLTRYEGQIKDVLV----- 273
65	Query: 269 GGYNSSTDTSNAVYYARSFVNPDPGKLGATGMTTQNTVEIDGVKNVLLIPSLTVKNRGG 328 + + ++A++YYAR VNPAG L MT Q +++ VKNVL IP + + G Sbjct: 274 -----TPEKVNDAIFYARFEVFNPNGLRLDMTAQVHQLTVDVKNVLTIFLSALGDGPV 328
	Query: 329 KAFVRV-LGADGKAVEREIRTGMKDSMNTVEKSGLKEGDKVVISSE 372 +V L +G+ ERE+ G ++ + E+ GL+ GD+VVI E Sbjct: 329 DNRYKVKLLRNGETREREVITGARNDTVEIVKGLGAGDEVVIGE 373

Based on this analysis, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF85-1 (40.4kDa) was cloned in the pGex vectors and expressed in *E.coli*, as described above.

The products of protein expression and purification were analyzed by SDS-PAGE. Figure 19A

- 5 shows the results of affinity purification of the GST-fusion protein. Purified GST-fusion protein was used to immunise mice, whose sera were used for Western blot (Figure 19B), FACS analysis (Figure 19C), and ELISA (positive result). These experiments confirm that ORF85-1 is a surface-exposed protein, and that it is a useful immunogen.

Example 92

- 10 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 773>:

```

1  ..ATTCCCGCCA CGATGACATT TGAACGACG CGCAATGCTT ACAAAATCGT
51  TTCCGACGATT AAAGTGCCGC TATACAAATAT CCGTTTCGAG TCCGGCGGTA
101  CGGTTGTGCGG CAATAGCCCTG CACCTTACCT ACTATAGAGA CATACGCGTA
151  GGCAAACTGT ATGCGGAAGc CAAATTCGCC GACGgGACGG TAACTTACGG
201  CAAAGCGGGC GAGAGCAAAA CGAGCAAAg CCCAAGGCT ATGGATTTGT
251  TCACGCTTGC CTGGGAGTTG GCGGCAAAAT ACGCGAAACT CCCCCGGGG
301  CTGAAATCA CCAACGGCAA AAAACTTTAT TCCTCGGG GTTTGAATAA
351  GCGCGGTACA GGAATAATCA GCTAGCGCGG CGTGGAAACC GAAGTCGTCA
401  AATATCGGT CCGCGCGGC GACATCGGG TATGTATTT cTTCGACCG
451  TCCTGAACA ATATTCCGC ACAATCGG TATACGAA cACGCAAAAC
501  CTATACGCTG AAATCAAAT CCGTGCAGAT CAACGCCAG CGACGCAAC
551  GTAA
  
```

This corresponds to the amino acid sequence <SEQ ID 774; ORF120>:

```

1  ..IPAIMTFERS GNAYKIVSTI KVPLYNIRFE SGGTVVGNLT HPTYYRDIRR
25  GKLYAEAKFA DGSVYTKAG ESKTEQSPKA MDLFTLAWQL AANDAKLPPG
101  LKITNGKKLY SVGGLNKAGT GRYSIGGVET EVVKYRVRRG DDAVMYFFAP
151  SLNNIPAQIG YTDGKTYTL KLSVQINGQ AAKP*
  
```

Further work revealed the complete nucleotide sequence <SEQ ID 775>:

```

1  ATGATGAAGA CTTTAAAAA TATATTTTC GCGCCATT TTGTCGCCGC
30  CCGCGCGTGC GCGTATGCGG CAGGGCTGCC CCAATCCGCC GTGCTGCACT
101  ATTCGCGCAG CTACGCGATT CCGCGCAGCA TGACATTTGA ACGCAGCGGC
151  AATCCTTACA AATGCTTTC GCGGATTAAG GTCCGCTAT ACATATTCOG
201  TTTCGAGTCC GCGGTACGG TTGTGCGCAA TACCTGCAC CTTACTACT
35  ATAGAGACTA ACGCAGGGGC AAACGTGATG CGGAGGCCAA ATTCCGCCAC
301  GGCAGCGTAA CTTACGCGCA AGCGGCGGAG AGCAAAACCG AGCAAAAGCCC
351  CAAGGCTATG GATTTGTTC ACGTTGCCCT GCAGTTGCGC GCAAAATGACG
401  CGAAACTCCC CCGGGGCTG AAAATCACCA ACGGCAAAA ACTTTATTTC
451  GTCGGCGGTT TGAATAAGGC GGGTACAGGA AAATACAGCA TAGGCGGCGT
501  GGAACCGGAA GTCGTCAAAT ATCGGGTGGC GCGCGGCGAC GATGCGGTAA
40  TGTATTTCTT CGCACGCTCC CTGAACATA TTCGCGGCA AATCGGCTAT
601  ACCGACGACG GCAAAACCTA TACGCTGAAA CTCAAATCGG TGCAGATCAA
651  CGGCCAGGCA GCCAAACCGT AA
  
```

This corresponds to the amino acid sequence <SEQ ID 776; ORF120-1>:

```

1  MMEKFNIFS AAILSAALPC AYAAGLPOSA VLHYSGSYGI PATMTFERSG
45  NAYKIVSTIK VPLYNIRFES GGTVVGNLT LH PTYRDIRG KLYAEAKFAD
101  GSVYTKAGE SKTEQSPKAM DLFTLAWQL ANDAKLPGL KITNGKKLYS
151  VGLNKGAGT KYSIGGVETE VVKYRVRRG DAVMYFFAP LNIPAQIGY
201  TDDGKTYTLK LKSVQINGQA AKP*
  
```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF120 shows 92.4% identity over a 184aa overlap with an ORF (ORF120a) from strain A of *N.meningitidis*:

```

5      orf120.pep      10      20      30      40      50      60      70      80      90
      orf120a      10      20      30      40      50      60      70      80      90
      SAAILSAALPCAYAAGLFXSAVLHYSGSYGIPATXXXXXXXNAXKIVSTIKVPLYNIRFE
      10      20      30      40      50      60      70      80      90
10     orf120.pep      100     110     120     130     140     150
      orf120a      100     110     120     130     140     150
      SGGTVVGNLHPTYYRDIRRGKLYAEAKFADGVSVTYKGAGESKTEQSPKAMDLFITLAWQL
      100     110     120     130     140     150
15     orf120.pep      160     170     180     190     200     210     220
      orf120a      160     170     180     190     200     210     220
      SGGTVVGNLHPTYYRDIRRGKLYAEAKFADGVSVTYKGAXXXXXXQSPKAMDLFITLAWQL
      160     170     180     190     200     210     220
20     orf120.pep      230     240     250     260     270     280     290     300
      orf120a      230     240     250     260     270     280     290     300
      AANDAKLPFGLKITNGKKLYSVGGLNKAGTGYKYSIGGVETEYVVKYRVRRGDDAVMYFFAP
      230     240     250     260     270     280     290     300
25     orf120.pep      310     320     330     340     350     360     370     380
      orf120a      310     320     330     340     350     360     370     380
      SLNNIPAQIGYTDGKTYTLKLSVQINGQAQKFX
      310     320     330     340     350     360     370     380
      SLNNIPAQIGYTDGKTYTLKLSVQINGQAQKFX
      310     320     330     340     350     360     370     380

```

The complete length ORF120a nucleotide sequence <SEQ ID 777> is:

```

30     1  ATGATGAAGA  CTTTAAAAA  TATATTTC  GCCGCCATT  TGTCCGCCG
      51  CCTGCCGTGC  GCGTATGCG  CAGGGCTGC  CNAATCGCC  GTGCTGCACT
      101 ATTCCGGCAG  CTACGGCAT  CCGCGCACNA  NNANNTNNG  ACNNNGNHC
      151 AATGCTTNCA  AAATCGTTTC  GACGATTAAA  GTGCCGTAT  ACATATCCG
      201 TTTCGAGTCC  GCGCGTACG  TTGTCGCAG  TACCTGTAC  CTTACTACT
      251 ATAGAGACAT  ACGCAGCGG  TTGTCGCAG  TACCTGTAC  CTTACTACT
      301 GGCAGCGTAA  CTTACGGCA  AGCGGNNNN  ANCNNNNNN  NGCAAAGCCC
      351 CAAGCGTATG  GATTGTCTCA  CGCTTGCNTG  CGAGTTGCC  GCAAATGACG
      401 CGAAATCCCC  CCGCGGGCTG  AAAATCACCA  ACGGCAAAA  ACTTTATTC
      451 GTCGGCGGTT  TGAATAAGGC  GGGTACAGGA  AAATACAGCA  TAGGCGGCGT
      501 GGAACCGCAA  GTCGTCAAT  ATCGGGTGG  GCGCGGCGAC  GATCGGGTAA
      551 TGTATTCTTT  CGCACCGTCC  CTGAACAATA  TTCGGGCACA  AATCGGCTAT
      601 ACCGACGACG  GCAAACCTA  TACGCTGAAA  CTCAAATCG  TGCAGATCAA
      651 CGGCCAGGCA  GCCAACCGT  AA

```

This encodes a protein having amino acid sequence <SEQ ID 778>:

```

45     1  MMKTFKNIFS  AAILSAALPC  AYAAGLFXSA  VLHYSGSYGI  PATXXXXXXX
      51  NAXKIVSTIK  VPLYNIRFES  GGTVVGNLH  PTYYRDIRRG  KLYAEAKFAD
      101 GSVTYGKAXX  XXXXQSPKAM  DLFTLAWQLA  ANDAKLPFGL  KITNGKKLYS
      151 VGLINKAGTG  KYSIGGVETE  VVKYRVRRGD  DAVMYFFAPF  LNNIPAQIGY
      201 TDDGKTYTLK  LKSVQINGQA  AKP*

```

ORF120a and ORF120-1 show 93.3% identity in 223 aa overlap:

```

50     orf120a.pep      10      20      30      40      50      60
      orf120-1      10      20      30      40      50      60
      MMKTFKNIFSAAILSAALPCAYAAGLFXSAVLHYSGSYGIPATXXXXXXXNAXKIVSTIK
      10      20      30      40      50      60
55     orf120a.pep      70      80      90      100     110     120
      VPLYNIRFESGGTVVGNLHPTYYRDIRRGKLYAEAKFADGVSVTYKGAXXXXXXQSPKAM
      70      80      90      100     110     120

```

5	orf120-1	VPLYNIRFESGGTVVGNLT	LHPTYYRDIRRGKLYAEAKFADG	SVTYKGAGESKTEQSPKAM	
		70	80	90	100
		110	120		
10	orf120a.pep	DLFTLAWQLAANDAKLP	PGLKITNGKKLYSVGLNKAGT	GKYSIGGVETE	VVKYRVRGDD
	orf120-1	DLFTLAWQLAANDAKLP	PGLKITNGKKLYSVGLNKAGT	GKYSIGGVETE	VVKYRVRGDD
		130	140	150	160
15	orf120a.pep	DAVMYFFAPSLNNIP	QAIGYTDGKTYTLKLKSVQING	QAAPFX	
	orf120-1	DAVMYFFAPSLNNIP	QAIGYTDGKTYTLKLKSVQING	QAAPFX	
		190	200	210	220

Homology with a predicted ORF from *N.gonorrhoeae*

ORF120 shows 97.8% identity over 184 aa overlap with a predicted ORF (ORF120ng) from *N.gonorrhoeae*:

20	orf120.pep	IPATMTFERSGNAYKIVSTIKVPLYNIRFE	30	
	orf120ng	SAAILSAAALPCAYAARLPQSAVLHYSGGYGIPTMTFERSGNAYKIVSTIKVPLYNIRFE	69	
25	orf120.pep	SGGTVVGNLTLHPTYYRDIRRGKLYAEAKFADGSVTYKGAGESKTEQSPKAMDLTFLAWQL	90	
	orf120ng	SGGTVVGNLTLHPAYYKDIRRGKLYAEAKFADGSVTYKGAGESKTEQSPKAMDLTFLAWQL	129	
30	orf120.pep	AANDAKLPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE	VVKYRVRGDDAVMYFFAP	150
	orf120ng	AANDAKLPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE	VVKYRVRGDDTVYFFAP	189
35	orf120.pep	SLNNIPAQIGYTDGKTYTLKLKSVQINGQAAP	184	
	orf120ng	SLNNIPAQIGYTDGKTYTLKLKSVQINGQAAP	223	

The complete length ORF120ng nucleotide sequence <SEQ ID 779> is:

35	1	ATGATGAAGA	CTTTTAAAAA	TATATTTTCC	GCCGCCATT	TGTCGCCGCCG
	51	CCTGCCGTGC	GCGTATGCGG	CAAGGCTACC	CCAATCCGCC	GTGCTGCATCT
40	101	ATTCGCCGAG	CTACGGCATT	CCCGCCACGA	TGACATTGGA	ACGCACGGCG
	151	AATGCTTACA	AAATCGTTTC	GACGATTAAA	GTGCCGCTAT	ACAATATCCG
45	201	TTTCGAATCC	GCGGTCACGG	TGTGCGGCAG	TACCCGTGAC	CCTGCTACT
	251	ATAAAGACAT	ACGCAGGGGC	AACTGTATG	CGGAAGCCAA	ATTGCGCGAC
50	301	GGCAGCGTAA	CCTACGGCAA	AGCGGCGGAG	AGCAAAACCG	AGCAAAACCG
	351	CAGAGCTATG	GATTTGTGCA	GCGTGCCTG	CAAAATGACG	CAAAATGACG
55	401	CGAACTCC	CCCGGTCG	AAAATACCA	ACGCGAAAA	ACTTTATTC
	451	GTGCGCGGCC	TGAATAAGGC	GCGTACGCGA	AAATACAGCA	TAGGCGGCGT
60	501	GGAAACCGAA	GTGCTCAAT	ATCGGTCGCG	GCGGCGGAC	GATACGCTAA
	551	CGTATTTCTT	CGCACGCTCC	CTGAACAATA	TTCGCGCAC	AATCGGCTAT
65	601	ACCGACGACG	GCAAAACCTA	TACGCTGAAG	CTCAATCGG	TGCAGATCAA
	651	CGGACAGGCC	GCAAAACCTG	AA		

This encodes a protein having amino acid sequence <SEQ ID 780>:

50	1	MMKTFKNIFS	AAILSAALPC	AYAAARLPQSA	VLHYSYSYGI	PATMTFERSG
	51	NAYKIVSTIK	VPLYNIRFES	GGTVVGNLT	PAYYKDIRRG	KLYAEAKFAD
55	101	GSVTYKGAGE	SKTEQSPKAM	DLFTLAWQLA	ANDAKLPGL	KITNGKKLYS
	151	VGLNKAGT	GKYSIGGVETE	VVKYRVRGDD	DTVYFFAPS	LNNIPQAIGY
60	201	TDGKTYTLK	LKSVQINGQA	AKP*		

In comparison with ORF120-1, ORF120ng shows 97.8% identity in 223 aa overlap:

50	orf120-1.pep	MMKTFKNIFS	AAILSAALPC	AYAAARLPQSA	VLHYSYSYGI	PATMTFERSGNAYKIVSTIK	
	orf120ng	MMKTFKNIFS	AAILSAALPC	AYAAARLPQSA	VLHYSYSYGI	PATMTFERSGNAYKIVSTIK	
60		10	20	30	40	50	60
		10	20	30	40	50	60

		70	80	90	100	110	120
5	orf120-1.pep	VPLYNIRFESGGTVVGNLHPPTYYRDIRRGKLYAEAKFADGSVYTGKAGESKTEQSPKAM					
	orf120ng	VPLYNIRFESGGTVVGNLHPAYYKDIRRGKLYAEAKFADGSVYTGKAGESKTEQSPKAM					
		70	80	90	100	110	120
10	orf120-1.pep	DLFTLAWQLAANDAKLPGLKITNGKKLYSVGLINKAGTGYKIGGVETEYVVKYRVRGD					
	orf120ng	DLFTLAWQLAANDAKLPGLKITNGKKLYSVGLINKAGTGYKIGGVETEYVVKYRVRGD					
		130	140	150	160	170	180
15	orf120-1.pep	DAVMYFFAPSLNNIPAQIGYTDGKTYTLKLSVQVNGQAAPX					
	orf120ng	DTVTYFFAPSLNNIPAQIGYTDGKTYTLKLSVQVNGQAAPX					
		190	200	210	220		

This analysis, including the presence of a putative leader sequence in the gonococcal protein suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 93

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 781>:

25	1	ATGTATCGGA	GGAAAGGGCG	GGGCATCAAG	CGGTGGATGG	GTGCCGGTGC
	51	.GCGTTTGCC	GCCTTGGTCT	GGCTGGTTTT	CGCGCTCGGC	GATACTTTGA
	101	CTCCGTTTGC	GTTTGGCGCG	GTGCTGGCGT	ATGTATTGGA	CCCTTTGGTC
	151	GAATGGTTGC	AGAAAAAGGG	TTTGAACCGT	GCATCCGCTT	CGATGCTCTG
	201	GATGGTGTTT	TCCTTGATTT	TGTTTGTGGC	ATTATTGTTG	ATTATCGTCC
30	251	CTATGCTGGT	CGGCGAGTTC	AACAATTTGG	CATCGCGCCT	GCCCCAATTA
	301	ATCGGTTTTA	TGCAGAACAC	GCTGCTGCGC	TGGTTGAAAA	ATACAATCGG
	351	CGGATATGTG	GAAATCGATC	AGGCATCTAT	TATTGGCTGG	CTTCAGGGCG
	401	ATACGGGAGA	GTTGAGCAAC	GCGCTTAAGG	CGTGGTTTCC	CGTTTGGATG
	451	AGGCAGGGCG	GCAATATT..			

This corresponds to the amino acid sequence <SEQ ID 782; ORF121>:

35	1	MYRRKGRGIK	PWMGAGXAF	ALVVLVFLAG	DTLTPFAVAA	VLAVLDPLV
	51	EWLQKKGLNR	ASASMSVMVF	SLIILLALLL	IIVPMLVGQF	NNLASRLPOL
	101	IGFMQNTLLP	WLKNTIGGYV	EIDQASIIAW	LQAITGELSN	ALKAWFFVLM
	151	RQGGNI..				

Further work revealed the complete nucleotide sequence <SEQ ID 783>:

40	1	ATGTATCGGA	GGAAAGGGCG	GGGCATCAAG	CGGTGGATGG	GTGCCGGTGC
	51	GCGGTTTGCC	GCCTTGGTCT	GGCTGGTTTT	CGCGCTCGGC	GATACTTTGA
	101	CTCCGTTTGC	GTTTGGCGCG	GTGCTGGCGT	ATGTATTGGA	CCCTTTGGTC
	151	GAATGGTTGC	AGAAAAAGGG	TTTGAACCGT	GCATCCGCTT	CGATGCTCTG
45	201	GATGGTGTTT	TCCTTGATTT	TGTTTGTGGC	ATTATTGTTG	ATTATCGTCC
	251	CTATGCTGGT	CGGCGAGTTC	AACAATTTGG	CATCGCGCCT	GCCCCAATTA
	301	ATCGGTTTTA	TGCAGAACAC	GCTGCTGCGC	TGGTTGAAAA	ATACAATCGG
	351	CGGATATGTG	GAAATCGATC	AGGCATCTAT	TATTGGCTGG	CTTCAGGGCG
	401	ATACGGGAGA	GTTGAGCAAC	GCGCTTAAGG	CGTGGTTTCC	CGTTTGGATG
50	451	CTTCGCTGCT	TACTATTTC	TGCTGGATTC	GACACCGTGC	TGCTGCTGCC
	501	TTGCCAAACT	GGTTCGAGG	CGTTTTCGCG	GTGCTTATAC	CGCATTACCA
	551	GGCAATTTGA	ACGAGGTATT	GGCGGAATTT	TTGCGCGGGC	AGCTTCTGGT
	601	AATGCTGATT	ATGGGCTTGG	TTTACGGTTT	GGGATTTGGT	CTGGTCGGGC
	651	TGGATTCGGG	GTTTGCCATC	GGTATGCTTG	CCGGTATTTT	GGTGTGTTGC
55	701	CCTTATCTCG	GGGCGTTTAC	GGGATTTGCT	CTTGCCACCG	TCGCGCGCCT
	751	GCTCCAGTTC	GGTTCGTGGA	ACGGCATCCT	ATCGGTTTGG	GCGGTTTTTG
	801	CCGTAGGACA	GTTTCTCGAA	AGTTTTTCCA	TTACGCCGAA	AATCGTGGGA
	851					

901 GACCGTATCG GCCTGTCGCC GTTTTGGGTT ATCTTTTCGC TGATGGCGTT
 951 CCGGCACTGT ATGGGCTTTG TCGGAATGTT GCGCGGATG CTTTGGCCGC
 1001 CCGTAACCTT GGTCTTGCTT CCGAGGGCG TGCAGAAATA TTTTGGCCGC
 1051 AGTTTTTACC GGGCAGGTA G

5 This corresponds to the amino acid sequence <SEQ ID 784; ORF121-1>:

1 MYRRKGRGIK PWMGAGAAFA ALVWLVFALG DTLTPFAVAAL VLAYVLDPLV
 51 EWLQKKGLNR ASASMSVMVF SLILLALLL IIVPMLVGQF NNLASRLPQL
 101 IGFMONTLLE WLKNTIGGVY EIDQASIIAW LQAHTGELSN ALKAWPPLVM
 151 RQGGNIVSST GNLLLLPLLL YFLLDWQRW SCGIAKLVPR RFAGAYTRIT
 201 GNINEVLGEF LRQQLVLMIL MGLVYGLGLV LVGLDSGFAI GMLAGILVEF
 251 FYLGAFTGLL LATVAALLQF GSWNGILSVN AVFAVGQFLE SFPTTPKTVG
 301 DRTGLSPFWV IFSLMAFGQL MGFVGMLAGL PLAAVTLVLL REGVQKYFAG
 351 SFYRGR*

Computer analysis of this amino acid sequence gave the following results:

15 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF121 shows 98.7% identity over a 156aa overlap with an ORF (ORF121a) from strain A of *N.*

meningitidis:

		10	20	30	40	50	60
20	orf121.pep	MYRRKGRGIK	PWMGAGXAF	ALVWLVFALG	DTLTPFAVA	AVLAYVLDPL	VEWLQKKGLNR
	orf121a	MYRRKGRGIK	PWMDAGAAFA	ALVWLVFALG	DTLTPFAVA	AVLAYVLDPL	VEWLQKKGLNR
		10	20	30	40	50	60
		70	80	90	100	110	120
25	orf121.pep	ASASMSVMV	FSLLILLALL	LIIVPMLVG	QFNNLASRL	PQLIGFMONT	LLEPWLKNTIG
	orf121a	ASASMSVMV	FSLLILLALL	LIIVPMLVG	QFNNLASRL	PQLIGFMONT	LLEPWLKNTIG
		70	80	90	100	110	120
		130	140	150			
30	orf121.pep	EIDQASIIAW	LQAHTGELS	NALKAWFP	VLMRQGGNI		
	orf121a	EIDQASIIAW	LQAHTGELS	NALKAWFP	VLMRQGGNI	SSIGNLLLL	PLLLYYFLLD
		130	140	150	160	170	180
35	orf121a	SCGIAKLVPR	RFAGAYTRIT	GNINEVLGE	FLRGQLVLM	IMGLVYGLG	LVLVGLDSGFAI
		190	200	210	220	230	240

The complete length ORF121a nucleotide sequence <SEQ ID 785> is:

40 1 ATGTATCGGA GGAAGGGCG GGCATCAAG CCGTGGATGG ATGCCGGTGC
 51 GGCCTTTGCC GCCTTGCTCT GCCTGGTTT CCGCTCGGC GATACTTTGA
 101 CTCGGTTTGC GTTTCGGCGG GTGCTGGCGT ATGTATTGGA CCCTTTGGTC
 151 GAATGGTTGC AGAAAAAGGG TTTGAACCGT GCATCGCCTT CGATGCTGCT
 201 GATGGGTGTT TCCTTGATTT TGTATGTTGC ATTTATGCTG ATTTATGCTC
 251 CTATGCTGCT CGGCGATTG AACAAATTGG CATCGCGCCT GCCCAATTAA
 45 301 ATCCGTTTTA TCCAGAACAC GCTCCTGCCG TGGTGAARA ATACAATCCG
 351 CGGATATCTG GAATATCGATC AGGCATCTAT TATTGCGTGG CTTCAGCGCG
 401 ATACGGGCGA GTTGAGCAAC GCGCTTAAGG CGTGGTTTCC CGTTTGTATG
 451 AGGCAAGGCG GCAATATTGT CAGCAGTATC GGCAACCTGC TGCTGCTTCC
 501 CTGTGCTGTT TACTATTTCG TGCTGGATTG GCAGCGGTGG TCGTGCGGCA
 551 TTGCCAAACT GGTTCGAGGG CGTTTTCGCG GTGCTATATC GCGCATTACA
 601 GGCAATTTGA ACGAGGTATT GGGCGAATTT TTGCGCGGGC AGCTTCGTGT
 651 GATGCTGATT ATGGGTTTGG TTTACGGCTT GGGGTTGGTG TCGGTCCGGC
 701 TGGATTCGGG GTTTGCAATC GGATATGGTG CCGGATATTT GGTTTTTGTT
 751 CCTATTITGG CGCGCTTTAC AGGACTGCTG CTGGCAACCG TCGCGCGCTT
 801 GCTCCAGTTC GGTTCGTGGA ACGGCATCTT GCGCTGTTGG GCGGTTTTTG
 851 CCGTAGGACA GTTCTCGGAA AGTTTTTCA TTACGCCGAA AATCGGGGA
 901 GACCGTATCG GCCTGTCGCC GTTTTGGGTT ATCTTTTCGC TGATGGCGCT
 951 CCGGCACTGT ATGGGCTTTG TCGGAATGTT GCGCGGATG CTTTGGCCGC
 1001 CCGTAACCTT GGTCTTGCTT CCGAGGGCG TGCAGAAATA TTTTGGCCGC
 60 1051 AGTTTTTACC GGGCAGGTA G

This encodes a protein having amino acid sequence <SEQ ID 786>:

5

```

1 MYRRKGRGK PWMDAGAAFA ALVWLVAFG DLTLTFVAFAA VLAYVLDPLV
51 EWLQKKGLNR ASASMYMVF SLILLALL ILVPMVLGQF NNLASRLPLQ
101 IGFMTNTLLP WLKNTIGGGV EIDQASIIAW LQAHTGELSN ALKANFPVLM
151 RQGGNIIVSSI GNILLPLLL YYFLVDQWRW SCGLAKLVPR RFAGAYTRIT
201 GNLNEVLGEF LRGQLLPLL MGLVYGLGVR LVGLDGSFAI GMVAGILVVF
251 PYLGAFITGLL LATVAALLQF GSWNGILAVW AVFAVGQFLE SFFITPKIVG
301 DRIGLSPFWV IFSLMAFGQL MGFVGMAGL PLAAVTLLVL REGVKYFAG
351 SFYRGR*

```

10 ORF121a and ORF121-1 show 99.2% identity in 356 aa overlap:

15

```

10 20 30 40 50 60
orf121a.pep MYRRKGRGKIKFWMDAGAAFAALVWLVFALGDTLTFVAFAAVLAYVLDPLVPEWLQKKGLNR
|||
orf121-1 MYRRKGRGIKFWMDAGAAFAALVWLVFALGDTLTFVAFAAVLAYVLDPEVLWLQKKGLNR
|||
10 20 30 40 50 60
70 80 90 100 110 120
orf121a.pep ASASMSVMVFSLLILLALLLIIVFVLVQGGNNLASRLPOLIGFMTNTLLPWLKNTIGGGV
|||
orf121-1 ASASMSVMVFSLLILLALLLIIVFVLVQGGNNLASRLPOLIGFMTNTLLPWLKNTIGGGV
|||
70 80 90 100 110 120
20
orf121a.pep EIDQASIIAWLQAHTGELSNALKAWPVLVLRQGGNIIVSSIGNLLLPILLYVFLLDQWRW
|||
orf121-1 EIDQASIIAWLQAHTGELSNALKAWPVLVLRQGGNIIVSSIGNLLLPILLYVFLLDQWRW
|||
130 140 150 160 170 180
25
orf121a.pep EIDQASIIAWLQAHTGELSNALKAWPVLVLRQGGNIIVSSIGNLLLPILLYVFLLDQWRW
|||
orf121-1 EIDQASIIAWLQAHTGELSNALKAWPVLVLRQGGNIIVSSIGNLLLPILLYVFLLDQWRW
|||
130 140 150 160 170 180
30
orf121a.pep SCGLAKLVPRRFAGAYTRITGNLNEVLGEFLRGQLVLMIMGLVYGLGLVVLGLDGSFAI
|||
orf121-1 SCGLAKLVPRRFAGAYTRITGNLNEVLGEFLRGQLVLMIMGLVYGLGLVVLGLDGSFAI
|||
190 200 210 220 230 240
35
orf121a.pep GMVAGILVVFVYPYLGAFITGLL LATVAALLQF GSWNGILAVW AVFAVGQFLESFFITPKIVG
|||
orf121-1 GMLAGILVVFVYPYLGAFITGLL LATVAALLQF GSWNGILSV AVFAVGQFLESFFITPKIVG
|||
250 260 270 280 290 300
40
orf121a.pep DRIGLSPFWVIFS LMAFGQLMGFVGMAGLPLAAVTLLVLREGVKYFAGSFYRGRX
|||
orf121-1 DRIGLSPFWVIFS LMAFGQMGFVGMAGLPLAAVTLLVLREGVKYFAGSFYRGRX
|||
310 320 330 340 350
45

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF121 shows 97.4% identity over a 156 aa overlap with a predicted ORF (ORF121ng) from *N. gonorrhoeae*;

50	orf121.pep	MYRRKGRGIGKFWMGAGAAPALVWLVFALGDTLPFAVAAVLAYVLDPLVWLQKKGLNR	60
	orf121.ng		
	orf121.ng	MYRRKGRGIGKFWMGAGAAPALVWLVFALGDTLPFAVAAVLAYVLDPLVWLQKKGLNR	60
55	orf121.pep	ASASMSVMVFSLLLLLLLTIVPMLVGQFNNLASRLPQLIGFMQNTLPLWLKNTIGGYV	120
	orf121.ng	ASASMSVMVFSLLLLLLLTIVPMLVGQFNNLASRLPQLIGFMQNTLPLWLKNTIGGYV	
	orf121.pep	EIDQASITAWLQAGHTELSNALKAFVPLMKGQNGIT	156
60	orf121.ng	EIDQASITAWFQAGHTELSNALKAFVPLMKGQNGIVSTIGNLLPLLLYYFLLDQWHRW	180

An ORF121ng nucleotide sequence <SEQ ID 787> was predicted to encode a protein having amino acid sequence <SEQ ID 788>:

```

1 MYRRKGRGIK PWMGAGAAFA ALVWLVLVYALG DTLTPFAVAA VLAYVLDELV
51 EWLQKKGLNR ASASMSVMVF SLILLALLL IIVPMLVGQF NNLASRLPQL
101 IGFMQNTLLP WLKNTIGGYV EIDQASIIAW FQAHTGELSN ALKAWFPVLM
151 KQGGNIVSTI GNLLLPPLLL YYFLLDWQRW SCGIKLVPR RFAGAYTRIT
201 GNLKVVWKF LRGQLLGETE RGAIVVCRVRG ECWEGGGARS RPSDDGWPRW
251 GGG*

```

Further work revealed the following gonococcal DNA sequence <SEQ ID 789>:

```

1 ATGTATCGGA GAAAAGGACG GGGCATCAAG CCGTGGATGG GTGCCGGCGC
51 GCGCTTTGCC GCCTTGGTCT GGCTGGTTTA CCGGCTCGGC GATACCTTTGA
101 CTCGCTTTGC GGTTCGGCGC GTGCTGGCGT ATGTGTGGGA CCCTTTGGTC
151 GANTGTTTGC AGAAAAAGGG TTGTAACCGT GCATCCGCTT CGATGCTCGT
201 GATGGTGT TTCTTGATTT TGTGTGTGGC ATTATPTGTT ATTATTGTCC
251 CTATGCTGGT CGGGCAGTTC AATAAATTGG CATCTCGCCT GCCCAATTA
301 ATCGGTTTTA TGCAGAACAC GCTGCTGCCG TGGTGTAAAA ATACAATCGG
351 CGGATATGTG GAATTCGATC AGGCATCTAT TATTGCTGG TTTCAAGCGC
401 ATACGGGCGA GTTGAGCAAC GCGCTTAAGG CGTGGTTTCC CGTTTTGATG
451 AAACAGGGCG GCAATATTGT CAGCAGTATC GCACACCTCG TGCTGCCGCC
501 CTGCTGCTCT TACTATTTCG TGCTGGATTG GCAGCGGCTGG TCGTGCGGCA
551 TCGCCAAACT GGTTCGAGGG CGTTTTCGCC GTGCTTATAC GCGCATACG
601 GGTAAATTGA ACGAGGTATT GGGCGAATTT TTGCGCGGTC AGCTTCTGTT
651 GATGCTGATT ATGGGCTTGG TTACGGT TTTGATGATG CTAGTCGAGC
701 TGGATTTCGG ATTTCGCATC GGTATGTTTG CCGGATTTT GGTGTTTGC
751 CCGTATTTCG GTGCGTTTAC GGGATTTCGT CTTCGACGTG TTGACGCTG
801 GCTCCAGTTC GTTCTCGGGA ACGGAATCTT GCGTGTTCGG GCGGTTTTTG
851 CCGTCGGTCA GTTCTCGGAA AGTTTTTTC TACGCGGAAA AATGCTAGGA
901 GACCGATCG GCTGTGCGCC GTTTTGGGTT ACTCTTTTCG TGATGGCGTT
951 CGGAGAGCTG ATGGGCTTTG TCGGAATGTT GGCGGGATTG CTTTGGCGCG
1001 CCGTAACCTT GCTCTTGTCT CGCGAGGCGC CGCAGAAATA TTTTCCGGCG
1051 AGTTTTTACC GGGCGAGSTA G

```

This corresponds to the amino acid sequence <SEQ ID 790; ORF121ng-1>:

```

1 MYRRKGRGIK PWMGAGAAFA ALVWLVLVYALG DTLTPFAVAA VLAYVLDELV
51 EWLQKKGLNR ASASMSVMVF SLILLALLL IIVPMLVGQF NNLASRLPQL
101 IGFMQNTLLP WLKNTIGGYV EIDQASIIAW FQAHTGELSN ALKAWFPVLM
151 KQGGNIVSSI GNLLLPPLLL YYFLLDWQRW SCGIKLVPR RFAGAYTRIT
201 GNLNEVLGEF LRGQLLVMLI MGLVYGLGLM LVGLDSGFAI GMVAGILFEV
251 PYLGAFTGLL LATVAALLQF GSWNGILAVW AVFAVGQFLE SEFTTKPIVG
301 DRIGLSPFWV IFSLMAGFEL MGFVGNLAGL PLAAVTLVL REGAQRYFAG
351 SFYRGR*

```

ORF121ng-1 and ORF121-1 show 97.5% identity in 356 aa overlap:

		10	20	30	40	50	60
orf121-1.pep	MYRRKGRGIK	PWMGAGAAFA	ALVWLVLVYALG	DTLTPFAVAA	VLAYVLDELV	EWLQKKGLNR	
orf121ng-1	MYRRKGRGIK	PWMGAGAAFA	ALVWLVLVYALG	DTLTPFAVAA	VLAYVLDELV	EWLQKKGLNR	
		10	20	30	40	50	60
orf121-1.pep	ASASMSVMVF	SLILLALLL	IIVPMLVGQF	NNLASRLPQL	IGFMTNTLLP	WLKNTIGGYV	
orf121ng-1	ASASMSVMVF	SLILLALLL	IIVPMLVGQF	NNLASRLPQL	IGFMTNTLLP	WLKNTIGGYV	
		70	80	90	100	110	120
orf121-1.pep	EIDQASIIAW	FQAHTGELSN	ALKAWFPVLM	KQGGNIVSSI	GNLLLPPLLL	YYFLLDWQRW	
orf121ng-1	EIDQASIIAW	FQAHTGELSN	ALKAWFPVLM	KQGGNIVSSI	GNLLLPPLLL	YYFLLDWQRW	
		130	140	150	160	170	180
orf121-1.pep	SCGIKLVPR	RFAGAYTRIT	GNLNEVLGEF	LRGQLLVMLI	IMGLVYGLGL	LVGLDSGFAI	
orf121ng-1	SCGIKLVPR	RFAGAYTRIT	GNLNEVLGEF	LRGQLLVMLI	IMGLVYGLGL	LVGLDSGFAI	
		190	200	210	220	230	240
orf121-1.pep	SGCIKLVPR	RFAGAYTRIT	GNLNEVLGEF	LRGQLLVMLI	IMGLVYGLGL	LVGLDSGFAI	
orf121ng-1	SGCIKLVPR	RFAGAYTRIT	GNLNEVLGEF	LRGQLLVMLI	IMGLVYGLGL	LVGLDSGFAI	

	190	200	210	220	230	240
5	orfl21ng-1	SCGIAKLVPRRFAGAYTRITGNLNEVLGEPLRGQLVLMILMGLVVGGLGIMLVGLDSDGFAI				
	orfl21-1.pep	GMLAGILVFVYPYLGAFTGLLLATVAALLQPGSWNGILSVWAVFAVGQFLSEFFITPKIVG				
10	orfl21ng-1	GMVAGILVFVYPYLGAFTGLLLATVAALLQPGSWNGILAVWAVFAVGQFLSEFFITPKIVG				
	orfl21-1.pep	DRIGLSPFWWIFSLMAFGQLMGFVGMLAGLPLAAVTLVLRLREGVQKYFAGSFYGRGX				
15	orfl21ng-1	DRIGLSPFWWIFSLMAFGELMGFVGMLAGLPLAAVTLVLRLREGAQKYFAGSFYGRGX				

In addition, ORF121ng-1 shows homology to a permease from *H. influenzae*:

```

sp|P43969|PERM_HAEIN PUTATIVE PERMEASE PERM HOMOLOG Length = 349
Score = 69.9 bits (168), Expect = 2e-11
Identities = 67/317 (21%), Positives = 120/317 (37%), Gaps = 7/317 (2%)

20 Query: 26 YVLAGDPLPFFVAAVLAAYLVLDPLVEWL--QKGLNRRASASMSVMVFSXXXXXXXXXXXXVP 84
    + YL GD + P + A VL+Y+L+ + +L Q R A++ + VF
  Sbjet: 32 IYFGDLIAPLLIALVLVSYLLETPINFLNQYLKCPRLMATLILFGSFIGLAAVFFVLVLP 91

25 Query: 85 MLVGQFNNLASRLPOLIGEMQNTLPLKNTIGGYE-IDQASIIAWFQAHGTGKSNALK 143
    + ML Q + L S LP + N W L N Y E ID + + + F + + +
  Sbjet: 92 MLWNQTLSLLSDLPAMF----NKSNEWLLNPKNYPELLIDYSMVDSIFRNSVREKILGFGE 147
  Query: 144 AWFVFLMKQGGNIVSSIGNXXXXXXXXXXXXXQWRWSCGIAKLPFRFAGAYRITGNL 203

30 + + + N+VS D G+++ +P+ A+ R + +
  Sbjet: 148 SAVKLSLASIMNVLNLSLGIYAFLVPLMMFMLEKDKSELQGVSRFLPKNNRLAFLRWK-EM 206

  Query: 204 NEVLGEFLRQGXKXXXXXXXXXXXXXXXXXSDGFAIGMVAGILVFVFPYKXXXXXXXXXX 263
    + + + G+ + + + + G+ V VFY
35  Sbjet: 207 QQQISNYHGKLLLEIIVLTITYIFLIFGLMYPLLLAFAGVSLVLFYIGAVIVTIEVA 266

  Query: 264 XXXXQIGFSWNGILAVNAVFAVGQFSLFFPIPKVIDRGIGLSPFWFIFLMAAGELMGF 323
    QFC + FAV Q L+ + P + + L P +I S++ PG L GF
  Sbjet: 267 LVMLPQFGISPIFFWYIIIAFAVSQDLQDGNLLVPYLFEAVNLLPLIIIIISVILFGGLWGF 326

40 Query: 324 VGLAGLPLAAVTLVLL 340
    G+ +PLA + ++
  Sbjet: 327 WGFVFAIEPLATLVKAVI 343

```

Based on this analysis, including the presence of a putative leader sequence and transmembrane domains in the two proteins, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 94

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 791>:

50	1	..	ACTGTCCTTT	CGCGCGGCGCT	GGCGTCAGAT	CATCATGAC	TGGTCATATAT
			TGGTTCCTTT	GGGAAACCGT	ATCAACAAAT	AGCGCGCATC	TTTAACATTTT
	101		TTTGCACGTC	CTCGCCGCGC	CGTTCAATGT	GTACCCAGCA	ATACCGCGCG
	151		CTGCGCGCTCT	ATGCGTTCCT	ATGCGCCGAG	TTCGCGGAT	TTTTCGTTGG
55	201		TTTTGCTTTT	GATTTGAGC	CAGAAATGT	CTATGCCAA	ATCGCGCGCG
	251		ATGTTGCGAC	CGATTTGCGC	AAGTGTGCGT	CGAGGTGTGG	TTTCTCTGGG
	301		ATATCGGCTC	CTGACACGTC	CGGACGCGTC	CGGACGCGTC	CGGACGCGTC
	351		TTTGTATAT	CGGACGCAA	AGACGCGGAT	CTTCCGCGAT	TTTGAACCTT
	401		CGCGCGCGTGT	CGGGGAATAT	TGTCGCGATA	TGCGCCAAAC	CTGCGCGCAC

451 GAGCAGCGCG TCGGTAACGG CGTGCAGCAG CGCATCGGCA TCGGAGTGTG
501 CGAGCAGCCC TTTTCAAAT GGGATTTCAA CTCCGCCAAG TATCAG..

This corresponds to the amino acid sequence <SEQ ID 792; ORF122>:

```

1  ..TAFSAALRLS PSXLVIFLSF GKPYQQTAAI LTFECTSCPP RSNAYQYRR
51 LRLVAFHPPE IAEFFVGFAF DVADNRNVAQ TGGDVGTHLR NVRRCCGFLC
101 NHRGRIIDRL PTLRINALIR LTKQDAWVR FELCGVGEM AADIAQTCTRI
151 EQRGVNGSEQP RIGIVGSEQP FFKWDFNSAK YQ..

```

Further work revealed the complete nucleotide sequence <SEQ ID 793>:

1	ATATCGTACT	GGGCAAGCAG	TCGCGGGATG	TTTGTGGAG	TAGATACCGG
51	GCCTTGTATG	TTTGTGGCGG	TTTCATCCAA	GGCTTCGTAG	AAAAGATGTT
101	TGGTCGAGCG	CGCGCGATG	CGGATATATT	CATTTTTCGTG	TACGAATCTG
151	ACTGGCTTTT	GCGGCGCAT	CGCTGTGAGT	TGCTGTTTGG	TGCTCATATT
201	TTTGTCCCTT	GGGAAACCGT	ATCAACAAC	AGCGCGCATC	TTAACATTTT
251	TTTTCGACGT	TCGCCCGCGG	CGTTCAAAAT	CGTACCAAGT	ATACCGCGCG
301	CTGCGCCCTT	ATGCTGTCCT	TCCGCGCGAG	ATAGCGCGAGT	TTTTCTGTTG
351	TTTGTGCTTT	GATGTGTGAC	CAGGAATATG	CTATGTCCAA	ATCGCGCGCG
401	ATGTGTGGAC	GCATTTTCGG	AATGTGCGGG	CGGAGTTTGG	GTTTCTGTGC
451	AATCAGCTG	TGATGACAT	TGACGCGCTG	CAACCCCTGC	GCCTGAACGC
501	TTTGTATCG	CGGCGCAAA	AGGACGCGCG	TGTCGCCCAT	TTTGAACCTT
551	GCGCGGGTGT	CGGCGAAATG	TGCGCGATG	TCGCCCAATC	TCGCCGCACC
601	GAGCAGCGCG	TCGGTACCG	TGCGCAGACG	CGCATCGGCA	TCGGAGTGTG
651	CSAGCAGCCC	TTTTTCAAA	GGGATTTCAA	TCGCGCCAGT	TATCAGCTTT
701	CTGCTTCGG	TCAGTGTGGT	GACATCGTAG	CCCTGTCCGA	TACGGATGTT
751	CGTCACTGTT	TGTGTTCTGT	A		

25 This corresponds to the amino acid sequence <SEQ ID 794; ORF122-1>:

1	ISYWASSPSSD	FLEVDYTFPLI	FLPLLPKASM	KKLMVEVPVM	PIYFSFGTNS
5	TAFSAAMRLS	SSCVVIFLASF	KQPKYQTAAI	LTFPCTSCPP	RSNAYQGYRR
10	LRLYALFVDE	TAEFFVFGVLF	DVDARNVYAQ	IGGVDVGTHLR	NVRREFGLCT
151	NHGRIDIDRL	PLRLNALTR	TR OKDAAYRI	FELCGGVGEM	AADTAQT CRT
201	EQRVNGGVQQ	RIGIVGSEQP	FFKWDFNSAK	YQLSAFGQLV	DIVALSPTDV
251	RRRLS*				

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF122 shows 94.0% identity over a 182aa overlap with an ORF (ORF122a) from strain A of *N*.

35 *meningitidis:*

[illegible]

210 220 230 240 250

The complete length ORF122a nucleotide sequence <SEQ ID 795> is:

5
10
15
20
25
30
35
40
45
50
55
60
65
70
75

```

1 ATATCATATT GGGCAAGCAG TTTCTGGAT TTTTGGAG TAGATACGCC
51 GCCTTTGATT TTTTGGCCG TCTTACCCAA GGCTTCGATG AAAAAGTTGA
101 TGGTCGAACC GGTACCGATG CCGATGTATT CGTTTTCGGG TACGAATTCG
151 ACTGCNTTTT CGGCGCGGAT GCGCTTGAGT TCGTCTGTG TCGTCATATT
201 TTTGTCCTTT GGGAAACCGT ATCAACAAAC AGCCGCCATC TTAACATTTT
251 TNNNACGTC CTGCCCGCGG CGTTCAAATC CTTACCAACA ATACCGCGCG
301 CTGCGACTCT ATGCGCTTCA TCGCGCCGAG ATACCCGAGT TTTTCGTTGG
351 TTTTGCTTTT GANGTTGACG CACGAAATGT CTATGCCCAA ATCGGGCGGG
401 ATGTTGGCAC GCATTTGCGG AATATGCGGG CGGAGTTTGG GTTTCGTGTG
451 AATCAGCGTC GTATCGACAT TGACCGCGCTG CCAACCCCTG GCCTGAAGCG
501 TTTGATACCG CGCAGCAAAA AGGACGCGGG TGTCGGCATC TTTGAACCTG
551 CGCGCGGTGT CGGGGAAATG GCTGCCGATA TCGCCCAACG CTGCCGACAC
601 GAGCAGCGCG TCGGTAACGG CGTGCAAGCG CGCATTTCAA TCGGAGTGTG
651 CGAGCAGCCC TTTTTCAAAT GGGATTGCAA CTGCGCCAAG TATCAGCTTT
701 CTGCTTCGGT TCAGTTGCTG GACATCGTAG CCCTGTCCGA TACGGATGTT
751 CGTCATCGTT TGTGTTCTCG A

```

This encodes a protein having amino acid sequence <SEQ ID 796>:

20
25

```

1 ISYWASSSLD FLEVDTAFLI FLPLLPKASM KKLMEVPVPM PMYSFSGTNS
51 TAFSAAMRLS SSCVVI FLSP GKPYQQTAAI LTFEFTSCPP RSNPYQQYRR
101 LRLYAFHAF ETEFFVGFAG XVDARNVYAG IGGDVGTHLR NMRREFGFLC
151 NHGRIDIDRL PTLRLNALIR RTQKDAAVRI FELCGVGEM AADIAQTCTR
201 EQRVGNVGQQ RIGIGVSEQP FFKWDFNSAK YQLSAFGQLV DIVALSDTDV
251 RHRLCS*

```

ORF122a and ORF122-1 show 96.9% identity in 256 aa overlap:

30
35
40
45
50
55

```

              10      20      30      40      50      60
orf122a.pep  ISYWASSSLD FLEVDTAFLI FLPLLPKASM KKLMEVPVPM PMYSFSGTNS TAFSAAMRLS
              |||||
orf122-1     ISYWASSSPD FLEVDTAFLI FLPLLPKASM KKLMEVPVPM PIYSFSGTNS TAFSAAMRLS
              10      20      30      40      50      60
              70      80      90     100     110     120
orf122a.pep  SSCVVI FLSP GKPYQQTAAI LTFEFTSCPP RSNPYQQYRR LRLYAFHAF ETEFFVGFAG
              |||||
orf122-1     SSCVVI FLSP GKPYQQTAAI LTFEFTSCPP RSNPYQQYRR LRLYAFHAF ETEFFVGFAG
              70      80      90     100     110     120
              130     140     150     160     170     180
orf122a.pep  XVDARNVYAG IGGDVGTHLR NMRREFGFLC NHGRIDIDRL PTLRLNALIR RTQKDAAVRI
              |||||
orf122-1     DVDAARNVYAG IGGDVGTHLR NVRRREFGFLC NHGRIDIDRL PTLRLNALIR RTQKDAAVRI
              130     140     150     160     170     180
              190     200     210     220     230     240
orf122a.pep  FELCGVGEMA ADIAQTCTRE QRVGNVGQQ RIGIGVSEQP FFKWDFNSAK YQLSAFGQLV
              |||||
orf122-1     FELCGVGEMA ADIAQTCTRE QRVGNVGQQ RIGIGVSEQP FFKWDFNSAK YQLSAFGQLV
              190     200     210     220     230     240
              250
orf122a.pep  DIVALSDTDV RHRLCSX
              |||||
orf122-1     DIVALSDTDV RHRLCSX
              250

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF122 shows 89.6% identity over a 182 aa overlap with a predicted ORF (ORF122ng) from *N.gonorrhoeae*:

	orf122.pep	TAFSAALRLSPSXLVIFLSFGKPYQQTAAI	30
	orf122ng	: : : :	80
5	orf122.pep	LTFFCTSCPPRSNAYQQYRRLRLYAFHPPEIAEFFVGFAGFDVDARNVYAQIGGDVGTHLR	90
	orf122ng	: : : : : : :	140
10	orf122.pep	NVRRCEGFLCNHGRIDIDLPLETRLNALIRRTQKDAAVRIFELCGVGEMAAIDIAQTCT	150
	orf122ng	: : : : : : :	200
	orf122.pep	EQRVNGVQQRIGIGVSEQPFKKWDFNSAKYQ	182
15	orf122ng	: : : :	256
	orf122ng	EQRVNGVQQRVGIRMFEPQPFKKWDFNSAKYQLSAFGLVDIVALSDTDIRHLCS	256

The complete length ORF122ng nucleotide sequence <SEQ ID 797> is:

	1	ATGTGCTACC	GGGCAAGCAG	TTCGCCGAGT	TTTTTGAGAG	TTGAAACGCG
	51	GCCTTGCATT	TTTTTACCGC	TTTTTCCCAA	GCCTTCGATG	AAGAATTTGA
	101	tggtCGAACC	GgtacCGATG	CCGATGTTAT	CGTTTTCGGG	TACGAATTCG
20	151	ACTGCTTTTT	CGCGCGCGAT	CGCCTtgAgT	TcgtctctgCG	TcgTCATATT
	201	TTTATctcttt	gGAAaccct	atcaAcaAac	agcgcccatCT	TTACATTTTT
	251	TTTGCAcGctc	ctggcgccgcg	cgttcaAATc	cgtaaccaGca	ataccgcgcgc
	301	ctgcgcctCT	AtgcCTTCCA	TCCGCCCGAG	ATAGCCGAGT	TTTTCGTfGG
	351	TTTTTGCTTT	GATatTGAOG	CACGAAATAT	CGatacCCAa	atcgcgcgCGG
25	401	ATGTTGGCAG	GCATTTCGGG	AATGTGCGGT	CGACAGTTGG	GTTTCTGTGC
	451	AATCACGCTC	GTATCGACAT	TGACCAcCTG	CCACCCCTGC	GCCTGAACGC
	501	TTTGATACGC	CGCACGCAAA	AGGACGCGGC	TGTCCGCATC	TTTGAACCTC
	551	GCGCGGTTGT	CGGGAAATG	GCTGCGCATG	TCCGCCAAAC	CTGCGCAACC
	601	GAGCAGCGcg	tcggttaaCGG	CGTGCAcGAG	cgcgTcgGca	TCCGAATGCC
30	651	CGAGCAGCCC	TTTTTCAAA	GGGATTTCAA	CTCCGCCAAG	TATCAGCTTT
	701	CTGCCCTCGG	TCAATTGGTG	GACATCGTAG	CCCTGTCCGA	TACGGATATT
	751	CGTCATCGTT	TGTGTTCTGT	A		

This encodes a protein having amino acid sequence <SEQ ID 798>:

	1	MSYRASSSPD	FLEVETAPLI	FLPLPKASM	KKLMVEVPFM	PMYSFSGTNS
35	51	TAFSAAMRLS	SSCVVIFLSF	GKPYQQTAAI	LTFFCTSWPP	RSNPYQYRR
	101	LRLYAFHPPE	IAEFFVGFAG	DIDARNIDTQ	IGGDVGTHLR	NVRCEFGFLC
	151	NHGRIDIDHL	FTLRNALIR	RTQKDAAVRI	FELCGGVGKM	AADVAQTCT
	201	EQRVNGVQQ	RVGIRMFEP	FFKWFNSAK	YQLSAFGLV	DIVALSDTDI
	251	RHRLCS*				

ORF122ng and ORF122-1 show 92.6% identity in 256 aa overlap:

		10	20	30	40	50	60
	orf122-1.pep	ISYWASSSPDFLEVD	TAPLI	IFLPLPKASM	KKLMVEVPFM	PMYSFSGTNS	TAFSAAMRLS
	orf122ng	: : : : : :					
45		10	20	30	40	50	60
	orf122-1.pep	SSCVVIFLSFGKPYQQTAAI	LTFFCTSCPPRSNAYQQYRRLRLYAFHPPEIAEFFVGFAG				
	orf122ng	: : : : : :					
50		70	80	90	100	110	120
	orf122-1.pep	SSCVVIFLSFGKPYQQTAAI	LTFFCTSCPPRSNAYQQYRRLRLYAFHPPEIAEFFVGFAG				
	orf122ng	: : : : : :					
		70	80	90	100	110	120
	orf122-1.pep	DVDARNVYAQIGGDVGT	HLNRVRE	FGFLCNHGRIDIDLPLETRLNALIRRTQKDAAVRI			
	orf122ng	: : : : : :					
		130	140	150	160	170	180
55		orf122-1.pep	DVDARNVYAQIGGDVGT	HLNRVRE	FGFLCNHGRIDIDLPLETRLNALIRRTQKDAAVRI		
	orf122ng	: : : : : :					
		130	140	150	160	170	180
	orf122-1.pep	FELCGGVGEMAAIDIAQTCT	EQRVNGVQQRIGIGVSEQPFKKWDFNSAKYQLSAFGLV				
	orf122ng	: : : : : :					
		190	200	210	220	230	240
60		orf122-1.pep	FELCGGVGEMAAIDIAQTCT	EQRVNGVQQRIGIGVSEQPFKKWDFNSAKYQLSAFGLV			
	orf122ng	: : : : : :					
		190	200	210	220	230	240
	orf122-1.pep	FELCGGVGEMAAIDIAQTCT	EQRVNGVQQRIGIGVSEQPFKKWDFNSAKYQLSAFGLV				
	orf122ng	: : : : : :					


```

                250
orf122-1.pep  DIVALSDTDVHRHRLCSX
5             |||||:||||:|||||
orf122ng      DIVALSDTDIRHRLCSX
                250

```

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 95

10 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 799>:

```

1  ..GCCGGCGCGA GTGCGAACAA CATTTCGCGG CGTTTTCGGG AACACCCGCT
51  CGCTGTCAGC GTTACCCCTGA TCGGCACGGT ACTTGCCTC APGCTGCCCG
101  TTACCGAATA TGAARACTTC CTGCTGCTTA TCGGCTCGGT ATTTCGCGCG
151  ATGGGGCGGA TTTTGATTGC CGACTTTTTC GTCTTGAAC GCGCTGA

```

15 This corresponds to the amino acid sequence <SEQ ID 800; ORF125>:

```

1  ..AGASANNISA RFAETPVAVS VTLIGTVLAV MLPVTEYENF LLLIGSVFAP
51  MGGFDCRLFR LETA*

```

Further work revealed the complete nucleotide sequence <SEQ ID 801>:

```

1  ATGTCGGGCA ATGCCTCCTC TCCTTCATCT TCCTCGGCCA TCGGGCTGAT
20  51  TTGGTTCGGG GCGGCGGTAT CGATTGCCGA AATCAGCAGC GGTACGCTGC
101  TTGCGCCTTT GGGCTGGCAG CGCGTCTGG CGGCTACTT TTTGGGTGAT
151  GCCGTCGCGG GCGCGCTGTT TTTTGGCGGG GCGTATATCG GCGCACTGAC
201  CGGACGCGAG TCGATGGAAA GCGTGCCTCT GTGCTTCGGG AACCGCGGTT
25  251  CAGTGCTGTT TTCCGTGGCG AATATGCTGC AACTGGCCGG CTGGACGCGG
301  GTGATGATTT ACGCGGCGCG AACGCTCAGC TCOCCTTTGG GCAAAGTGTT
35  351  GTGGGACGGC GAATCTTTTG TCTGTGGGGC ATTGGCAAC GCGCGCTGA
401  TTGTGCTGTG GCTGGTTTTC GCGCACGCA AACACGGCGG GCTGAARAC
451  GTTTCGATGC TGCTGATGCT GTTGGCGGTT CTGTGCTGA GTGCCGAGT
30  501  CTTTCCACAG CGAGGCGACA CGCGCGACA GCTTTCAGAC GCGATGAGT
551  TCGGAACGGC AGTCGAGCTG TCGCGCTGA TCGCGCTTC CTGGCTGCGG
601  CTGCGCCCGC ACTACACCGG CCACGCGCGC CGCCGCTTTC CGGCAACCTT
651  GACGCAACGC CTCGCTTACA CGCTGACCGG CTGCTGGATG TATGCTTGG
35  701  GTTGGGCAGC GCGCTTGTTC ACCGGAGAAA CCGACGTGGC AAAAATCCTG
751  CTGGGCGCAG GTTGGGTGCG GCGACGGCAT TTGGCGGTGC TCCTCTCCAC
801  CGTTACCACA ACGTTTCTCG ATGCTATTTC CGCGCGCGCG AGTGCGAACA
851  ACATTTCCGC GCGTTTTCGG GAACACCGC TCGCTGTCCG CGTTACCCCTG
901  ATCGGCAACG TACTTGGCCT CATGCTGCCC GTTACCGAAT ATGAAACTT
951  CCTGCTGCTT ATCGGCTCGG TATTGCGCC GATGGCGCGC GTTTTGATTG
40  1001  CCGACTTTTTC CGTCTTGAAA CGCGCTGAGG AGATTGAAGG CTTTGACTTT
1051  GCGGACTGGT TCTGTGGGCT TCGGGGCTTC ATCCCTACCC GCTTCTGCTG
1101  CTCGTCCGCG TGGGAAAGCA GCATCGGTCT GACCGCCCCC GTAATGTGCT
1151  CCGTTGCCAT TGCCACCGTA TCGGTACGCC TTTTCTTAA AAAACCCCAA
1201  TCTTTACAAA GGAACCCGTC ATGA

```

This corresponds to the amino acid sequence <SEQ ID 802; ORF125-1>:

```

1  MSGNASSPSS SSAIGLINFQ AAVSIAEIST GTLLAPLQWQ RGLAALLGH
45  51  AVGGALFPAA AYIGALTGRS SMESVRLSPG KRGSVLFPSVA NMLQLAGWTA
101  VMYYAGATVS SALGKVLWDG ESWFNWALAN GALIVLWLVF GARKTGGLKT
151  VSMMLMLLAV LWLSAEVFST AGSTAAQVSD GMSGTAVEL SAVMPLSWLP
201  LAADYTRHAR RPPAATLTAT LAYLTGCGWM YALGLAALF TGEDTVAKLL
50  251  LGAGLGAAGI LAVVLSTVTT TFLDAYSAGA SANNISARFA ETFFAVGVTL
301  IGTVLAVMLP VTEYENFLLL IGSVFAPMAA VLIADFFVLK RREEIEGFFD
351  AGLVLWLAGF ILYRPLSSG WESSIGLTAP VMSAVAIATV SVRLFKKTKQ
401  SLQRNPS*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF125 shows 76.5% identity over a 51aa overlap with an ORF (ORF125a) from strain A of *N. meningitidis*:

[illegible]

15 The ORF125a partial nucleotide sequence <SEQ ID 803> is:

	1	ATGTCGGCGGA	ATGCGCTCCTC	TGNTTCATCT	TCGGCGCGCCA	TGGCGCTGAT
	51	TTCGTTTCCGC	CGGCGGGTAT	CGATTCCGGA	CAATCAGCAGC	TGGGACTCGT
	101	TTCGCGCCTTT	GGGCTGGCAG	CTCGGCTCTGG	CNGCTCTGCT	TTTGGGTTCAT
20	151	CGCGTCGCGG	CGGCTGCTGT	TTCGTCGGCG	GGGTATATGCT	CGCCTATGAC
	201	CGGACGNCNC	TCGATGGAA	CGGCGCGCTC	GTCTGTCGGG	AAACCGGGTT
	251	CAGTGCCTGTT	TTCGTCGGCG	AAATACGCTGC	AATTCGGCGG	CTGGAACGGTT
	301	GTGATGATTT	ACGCGCGGCG	AACGGTCTGC	TCGGCTTTGG	CGAAAGCTGT
	351	GTGGGACGCG	GAACTTTTTC	TTCTGGTGGC	ATTGGCAAA	GGCGCGTGGT
25	401	TTTGGCTGTG	GCTGTTTTCG	GGGCGACGCA	AAACAGCGGG	GCTGAAACGG
	451	GTTTTCGATGT	TGCTGATGCT	TTGTGGCGGT	CTGTGGCTGA	GTGCCGAANT
	501	NTTTTCCACG	CAGCGGACGA	CCGCGCGNAN	GGTNACGAC	GGCATGAGTT
	551	TCGGAACCGC	GCTGAGCTG	TCGCGCGTTC	TGCGCTTTC	TTGCGTCGCG
	601	CTGCGCGCGC	ACTACGCTGC	CCAGCGCGCG	CGCGCGTTTG	CGGCACCCCT
	651	CAGGCGCAGC	CTCGCTCACA	CGCTGACGCG	CTGCTGAGT	TATGCTCTG
30	701	GTTTTCGCGC	GGGCTGCTTC	ACGCGGAGAA	CTGCTGCTGG	AAATATGCT
	751	GTTCGCGGCT	CTTTTTCCTG	ATGCTNACTC	TGTCGCGCTA	CTGCTCGAC
	801	GCTTACGAC	CTTTTTCCTG	ATGCTNACTC	CGCGCGCGTA	AGTGCACGTA
	851	ATATTTCGCG	CAACTTTTCG	GAAATACGNA	TCGCGCGTTC	CGTTCGCGTT
	901	GTGCGCACAC	TGCTTTCGGT	CTCTCTGGCG	GTATCCGAAT	ATGAAAACTT
35	951	CTGCTGTGCT	ATTGGGCTGT	TATTTGGCGC	GATGGGCGCG	TTTTTTGATTG
	1001	CCGACTTTTT	CGTCTTGAAA	CGGCTGTAGG	AGATTGAAAG	C..

This encodes a protein having the partial amino acid sequence <SEQ ID 804>:

```

1 MSGNASSXSS SAATGLIWFQ AVSIAIEIST KTLPLAPLWQ RGLAALLGH
51 AVGGALFFFA AYIAGLTGX SMESVRLSFG KRSLVFSVA NMLQLACGHT
101 VHYIYAGTAS SATGKVLWGD ESEFWVA GATVLYRVLV GARKTGKGLT
151 VMLMLMLLAV LMLSAKSTXV DMSATAXVD GMSFTQVLFV SAMPVLSGLT
201 LAADYTRHAR REPATLTAT LAYLTTCGM YALGLAAFL TGTDTAVKL
251 LSGAGLAGAI LVLVLTSTTT TFLDAYSAR SANNISAKLS EPIAVAVLW
301 VTGTLAVLLP VTETENVTLL TGSVFAPMAA VLIADFFVLK RREETEG..

```

45 ORF125a and ORF125-1 show 94.5% identity in 347 aa overlap:

		10	20	30	40	50	60
	orf125a.pep	MSGNASSXSSSAAGLWFGAAVSI	IAETISGTTLLAPL	GWQRLAALLLGH	AVGGALFFAA		
50	orf125-1	MSGNASSPSSSAAGLWFGAAVSI	IAETISGTTLLAPL	GWQRLAALLLGH	AVGGALFFAA		
		10	20	30	40	50	60
	orf125a.pep	70	80	90	100	110	120
55	orf125-1	AYIGALTGXSSMESVRLSP	FGKRGVLSVANNMLQ	LQAGTAVMIYAGATV	SSSALGKVLWDG		
		70	80	90	100	110	120
60	orf125a.pep	ESFVFWALANGALIVLWLV	FGARKTGGKLTYSMLL	MLLAVLWLSAEKFS	TAGSTAAKXVD		

	orf125-1	ESFVWVALANGALIVLWLVFGARKTGGKTVSMLMLLAVLWLSAEVFTAGSTAAQVSD	130	140	150	160	170	180
5	orf125a.pep	GMSFGTAVELSAVMPLSWLPLAADYTRHARRPFAATLTATLAYTLTGCMYALGLAALF	190	200	210	220	230	240
	orf125-1	GMSFGTAVELSAVMPLSWLPLAADYTRHARRPFAATLTATLAYTLTGCMYALGLAALF	190	200	210	220	230	240
10	orf125a.pep	TGETDVAKILLGAGLGAAGILAVLVSTVTTTFLDAYSGAGVSANNISAKLSEIPIAVAVAV	250	260	270	280	290	300
	orf125-1	TGETDVAKILLGAGLGAAGILAVLVSTVTTTFLDAYSGAGVSANNISARFAETPVAVGVTL	250	260	270	280	290	300
15	orf125a.pep	VGTLLAVLLPVTEYENFLLIGSVFAPMAAVLIADFFVLKRREEIEG	310	320	330	340		
	orf125-1	IGTVLAVMLPVTEYENFLLIGSVFAPMAAVLIADFFVLKRREEIEGFDFAGLVWLAFG	310	320	330	340	350	360

Homology with a predicted ORF from *N.gonorrhoeae*

ORF125 shows 86.2% identity over a 65aa overlap with a predicted ORF (ORF125ng) from

25	<i>N.gonorrhoeae</i> :								
	orf125.pep					AGASANNISARFAETPVAVSVTLIGTVLAV		30	
	orf125ng	KILLGAGLGITGILAVVLSTVTTTFLDTSYAGASANNISARFAEIPVAVGVTLIRTVLAV						308	
30	orf125.pep	MLPVTEYENFLLIGSVFAPM-GGDFCRLEFLETA				64			
	orf125ng	MLEVTEYKNFLLIRSVFGPMAGGDFCRLEFCLKTA				343			

An ORF125ng nucleotide sequence <SEQ ID 805> was predicted to encode a protein having amino acid sequence <SEQ ID 806>:

35	1	MSGNASSPSS	SAAIGLVWFG	AAVSTAEIST	GTLLAPLGWQ	RGLAALLLGH	
	51	AVGGALFFAA	AYIGALTGRS	SMESVRLSFG	KCGSVLFSVA	NMLQLAGWTA	
	101	VMIYVGATVS	SALGKVLWDG	ESFVWVALAN	GALIVLVLVF	GARRTGGKLT	
	151	VSMLLMLLAV	LWLSVEVFAS	SGTNAAPAVS	DGMTFGTAVE	LSAVMPLSWL	
	201	PLAADYTRQA	RRPFAATLTA	TLAYTLTGCV	MYALGLAAL	FTGETDVAKI	
40	251	LLGAGLGITG	ILAVVLSTVT	TFFLDTSYAG	ASANNISARF	AEIPVAVGV	
	301	LIRTVLAVML	PVTEYKNFLL	LIRSVFGPMA	GGDFCRLEFCL	KTA*	

Further work revealed the following gonococcal DNA sequence <SEQ ID 807>:

	1	ATGTCGGGCA	ATGCCCTCCTC	TCCTTCATCT	TCGCGCGCCA	TCGGGCTGGT	
45	51	TTGGTTTCGGC	CGGCGGGTAT	CGATCCCGGA	AATCAGCAG	GGTACGCTGC	
	101	TCGCCCCCTT	GGGCTGCAG	CGCGCTCGG	CGGCGCTGCT	TTGGGTATAT	
	151	CGCGTCGGCG	GCSCGCTGT	TTTTCGGCGG	CGGTATATAT	CGGCATGAC	
	201	CGAGCGCAGC	TCGATGGA	GTGTGCGCT	GTGCTTCGCG	AAATGCGGTT	
	251	CAGTGCCTGT	TTCCGTGGCG	AATATGCTGC	AACCTGGCGG	CTGGACGCGG	
	301	GTGATGATT	ACGTGCGCGC	AACGGTCAGC	TCCGCTTGG	GCAAGTGT	
50	351	GTGGGACGGC	GAATCCTTTG	TCTGGTGGCG	ATTGGCAAC	GGCGCACTGA	
	401	TGCTGCTGTG	CGTGTGTTTC	GGCGCAACGA	GAACGGCGCG	GCTGAAAC	
	451	GTTTCGATGC	TGCTGATGCT	GCTTGCCTGT	TTGTGGTTGA	CGCTCGAAGT	
	501	GTTTCGCTTCG	TCGCGGCACA	ACGCGCGCGC	CGCGCTTCA	GACGCGATGA	
	551	CCTTCGGAAC	GGCAGTCGAA	CTGTCCGCGC	TGATGCGCGT	TTCTTGCGGT	
55	601	CGCGTGGCGG	CGGACTACAC	CGCGCAAGCA	CGCGCGCGT	TTGCGGCAAC	
	651	CCTGACGGCA	ACGCTCGCTC	ATACGCTGAC	GGGCTGCTGT	ATGATGCTCT	
	701	TGGGTTTGGC	GGCGGCTCTG	TTTACCGGAG	AAACCGGAGT	GGGGAATATC	
	751	CTGTTGGGCG	CGGCGCTTGG	CATAACGGGC	ATTCTGGCAG	TGCTCTCTCT	
	801	CACCGTTTAC	ACACGCTTTC	TGATACCTA	TTCCGCGCGC	CGGAGTGGGA	
60	851	ACAACATTTTC	CGCGCGTTTT	CGGGAATATC	CGGTCGCTGT	CGGCGTTACG	

5 901 CTGATCGGCA CGGTGCTTGC CGTCATGCTG CCCGTTACCG AATATAAAAA
 951 CTTCCTGCTG CTTATCGGCT CGGTATTGCG GCCGATGGCG GCGGTTTTGA
 1001 TTGCGGCACTT TTTCGTCTTA AAGCGCGCTG AGGAGATTGA AGGCTTTGAC
 1051 TTGCGCGGAC TGGTTCCTGT GCTGGCAGCG TTCATCCTCT ACCGCTTCCT
 1101 GCTCTCGTCC GGTGGGAAA CAGCATCGG CTGACCGGCC CCGTAAATGT
 1151 CTGCGCTTGC CATTGCCACC GTATCGGTAC GCCTTTTCTT TAAAAAACC
 1201 CAATCTTTAC AAGGAACCC GTCATGA

This corresponds to the amino acid sequence <SEQ ID 808; ORF125ng-1>:

10 1 MSGNASSESS SAAIGLVWFG AAVSIAEIST GTLLAPLGWQ RGLAALLLGH
 51 AVGGALFFAA AYIGALTGRS SMESVRLSFG KCGSVLFSVA NMLQLAGWTA
 101 VMIVYGATVS SALGKVLWDG ESFVWALAN GALIVLWLVF GARRTGGLKT
 151 VSMLMLLLAV LWLSVEVFAS SGTNAAPAVS DGMTFGTAVE LSAMVPLSWL
 201 PLAADYTRQA RRPFAATLTA TLAYTLTGCV MYALGLAAAL FTGETDVAKI
 251 LLGAGLGITG ILAVVLSTVT TFLDFTYSAG ASANNISARF AEIIPVAVGT
 15 301 LIGTVLAVML PVTEYKNFLL LIGSVFAPMA AVLIADFFVL KRREEIEGFD
 351 FAGVLVWLAG FILYRFLSS GWESSIGLTA FVMSAVAIAT VSVRLFPPKT
 401 QSLQRNPS*

ORF125ng-1 and ORF125-1 show 95.1% identity in 408 aa overlap:

20	orf125-1.pep	MSGNASSESSSAAIGLVWFGAAVVSIAEISTGTLLAPLGWQRLAALLLGHAVGGALFFAA	10	20	30	40	50	60
	orf125ng-1	MSGNASSESSSAAIGLVWFGAAVVSIAEISTGTLLAPLGWQRLAALLLGHAVGGALFFAA	10	20	30	40	50	60
25	orf125-1.pep	AYIGALTGRSSMESVRLSFGKRGSVLFSVANMLQLAGWTAVMYIAGATVSSALGKVLWDG	70	80	90	100	110	120
	orf125ng-1	AYIGALTGRSSMESVRLSFGKRGSVLFSVANMLQLAGWTAVMYIAGATVSSALGKVLWDG	70	80	90	100	110	120
30	orf125-1.pep	ESFVWALANGALIVLWLVFGARKTGGLKTVSMMLLLAVLWLSAEVFSFAGSTAAQ-VS	130	140	150	160	170	179
	orf125ng-1	ESFVWALANGALIVLWLVFGARKTGGLKTVSMMLLLAVLWLSAEVFSFAGSTAAQ-VS	130	140	150	160	170	180
35	orf125-1.pep	DGMSFGTAVELSAVMPPLSWLPLAADYTRHARRPFAATLTATLAYTLTGCVMYALGLAAAL	180	190	200	210	220	230
	orf125ng-1	DGMTFGTAVELSAVMPPLSWLPLAADYTRHARRPFAATLTATLAYTLTGCVMYALGLAAAL	180	190	200	210	220	230
40	orf125-1.pep	FTGETDVAKILLGAGLGAAGILAVVLSTVTTTFLDAYSAGASANNISARFAETPVAVGVT	240	250	260	270	280	290
	orf125ng-1	FTGETDVAKILLGAGLGAAGILAVVLSTVTTTFLDAYSAGASANNISARFAETPVAVGVT	240	250	260	270	280	290
45	orf125-1.pep	LIGTVLAVMLPVTEYKNFLLIGSVFAPMAAVLIADFFVLKRREEIEGFDFAVLVWLAG	300	310	320	330	340	350
	orf125ng-1	LIGTVLAVMLPVTEYKNFLLIGSVFAPMAAVLIADFFVLKRREEIEGFDFAVLVWLAG	300	310	320	330	340	350
50	orf125-1.pep	FILYRFLSSGWESSIGLTA FVMSAVAIAT VSVRLFPPKTQSLQRNPSX	360	370	380	390	400	
	orf125ng-1	FILYRFLSSGWESSIGLTA FVMSAVAIAT VSVRLFPPKTQSLQRNPSX	360	370	380	390	400	

60 Based on this analysis, including the presence of putative leader sequence and transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and

N.gonorrhoeae, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 96

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 809>:

```

5      1  ATGACCCGTA  TCGCCATCCT  CGGCGGCGGC  CTCTCGGGAA  GGCTGACCGC
    51  GTTGACGCTT  CGAGACAAG  GTTATCAGAT  TGCACCTTTC  GATAAAGCT
   101  CGCGCGGGG  CGAGACAGCC  GCGCGCTATG  TAGCGCGCG  CATGCTCGCG
   201  CGAGAGCATC  CGCGTTTGGC  CGGCGATCCG  ATGCGTCTG  AACACGCACA
   301  CGATGATGCA  GGAAAACGGC  AGCCTGATTG  TATGGCAGCG  GCAGGACAAG
   351  TGACGAAATC  GTCGGTTGGC  GCGCGGACGA  CATCGCCGAA  CGCGAACCGC
   401  AACTCGGCGG  ACGTTTITTA  GACGGCATCT  ACCTGCCGAC  CGAAGC.CAG
   451  CTCGACGGGC  GGCAATTATA  GTCTGCACCT  GCGGACGCTT  TGGACGAAC
   501  GAACGTCGCC  TGCCATTGGG  AACACGAATG  GCTCCCGGAA  GCCTGCAAG..

```

This corresponds to the amino acid sequence <SEQ ID 810; ORF126>:

```

20      1  MTRIAILGGG  LSGRLTALQL  AEQYQIALF  DKSCRRGEHA  AAYVAAAMLA
    51  PAAXTVEATP  EVVRLGRQSI  PLWRGIRCLR  NHTMMQENG  SLIVWHGQDK
   101  PLSSEFVRHL  KRGGXTDDEI  VRWRADDIAE  REPQLGGRFX  DGIYLPTEQX
   151  LDGRQLXSAL  ADALDELNVP  CHWEHECVPE  ACK...

```

Further work revealed the complete nucleotide sequence <SEQ ID 811>:

```

25      1  ATGACCCGTA  TCGCCATCCT  CGGCGGCGGC  CTCTCGGGAA  GGCTGACCGC
    51  GTTGACGCTT  CGAGACAAG  GTTATCAGAT  TGCACCTTTC  GATAAAGCT
   101  CGCGCGGGG  CGAGACAGCC  GCGCGCTATG  TTAGCGCGCG  CATGCTCGCG
   151  CTGCGGCGGG  AAGCGGTGCA  AGCCACGCC  GAAGTGTCAC  GGCTGGGCGC
   201  CGAGAGCATC  CGCGTTTGGC  CGGCGATCCG  ATGCGTCTG  AACACGCACA
   251  CGATGATGCA  GGAAAACGGC  AGCCTGATTG  TATGGCAGCG  GCAGGACAAG
   301  CCATATATCCA  GCGAGTTTGT  CGGCCATCTC  AAACGCGGCG  CGGTAGCGGA
   351  TGACGAAATC  GTCGGTTGGC  GCGCGGACGA  CATCGCCGAA  CGCGAACCGC
   401  AACTCGGCGG  ACGTTTITTA  GACGGCATCT  ACCTGCCGAC  CGAAGGCCAG
   451  CTCGACGGGC  GGCAATATT  GTCTGCACCT  GCGGACGCTT  TGGACGAAC
   501  GAACGTCGCC  TGCCATTGGG  AACACGAATG  GCTCCCGGAA  GGCTGCAAG
   551  CCCAATACGA  CTGGGTGATC  GACTGCGCGG  GCTACGGCGC  AAAAACCGCG
   601  TGGAAACCAAT  CCCCAGAGCA  CACGAGACCC  TACGCGGGCA  TAGCGGGCGC
   651  AGTGGCGCGG  GTTTACACAC  CCGAATACAC  GCTCAACCGC  CCGTGCGGTC
   701  TGCTCCATCC  GCGGTATCCG  CTCTACATCG  CCGCGAAGA  AAACCCAGTC
   751  TTGCTCATG  GCGCGACCA  AATCGAAGC  GAAGACGAG  CCGCGGACG
   801  CGTGCTCAT  GGGTTGGAAC  TCTTTGCGG  ACTCTATGCC  ATCCACCGCG
   851  CCTTGGGCGA  AGCGACATC  CTCGAAATCG  CCACCGGCT  GCGCCCCAGC
   901  CTCACCAACC  ACAACCCGGA  AATCGGTAC  AACCGCGGCC  GACGCTGAT
   951  TGAATCAAC  GCGCTTTTCC  GCACGCGTTT  CATGATCTCC  CCGCGGTAA
  1001  CGCGCGCGC  CGCCAGATTG  GCAGTGGCAC  TGTTTGACGG  AAAAGACGGG
  1051  CCGGAACGCG  ATAAAGAAAG  CGGTTTGGCG  TATATCCGAA  GACAAGATTA
  1101  A

```

This corresponds to the amino acid sequence <SEQ ID 812; ORF126-1>:

```

50      1  MTRIAILGGG  LSGRLTALQL  AEQYQIALF  DKSCRRGEHA  AAYVAAAMLA
    51  PAAXTVEATP  EVVRLGRQSI  PLWRGIRCLR  NHTMMQENG  SLIVWHGQDK
   101  PLSSEFVRHL  KRGGVADDEI  VRWRADDIAE  REPQLGGRFX  DGIYLPTEQX
   151  LDGRQLXSAL  ADALDELNVP  CHWEHECVPE  GLAQYDNL  DCRGYGARTA
   201  WNQSPHSTST  LRIGIRGEVAR  VYTFEITLNR  PVRLHPHPRY  LYIAPKRNVT
   251  FVIGATQIES  ESQAPASVRS  GLELLSALYA  IHPAFGEADI  LEIATGLRPT
   301  LNHNEIRY  NRRLIEIN  GLFRHGFMS  FAVTAAARL  AVAFDPGKDA
   351  PERDKESGLA  YIRRPQ*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF126 shows 90.0% identity over a 180aa overlap with an ORF (ORF126a) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
5	orf126.pep	MTRIAILGGGLSGRLTALQLAEQGYQIALFDKSCRRGEHAAAYVAAAMLAPAAKTVETAP					
	orf126a	MTRIAILGGGLSGRLTALQLAEQGYQIALFDKSCRRGEHAAAYVAAAMLAPAAEAVEATP					
		10	20	30	40	50	60
10	orf126.pep	EVVRLGRQSIPLWRGIRCLRNTHTMQENCSLIIVWHGQDKPLSSEFVRHLKRGGXDTDEI					
	orf126a	EVVRLGRQSIPLWRGIRCLRNTHTMQENCSLIIVWHGQDKPLSSEFVRHLKRGGVADDXI					
		70	80	90	100	110	120
15	orf126.pep	VWRWADDIAEREPLQGGRFSDGIYLPTEQLDGRQLXSALADALDELNVCHWEHECVPE					
	orf126a	VWRWADDIAEREPLQGGRFSDGIYLPTEQLDGRQLXSALADALDELNVCHWEHECAFE					
		130	140	150	160	170	180
20		130	140	150	160	170	180

The complete length ORF126a nucleotide sequence <SEQ ID 813> is:

	1	ATGACCCGTA	TCGCCATCCT	CGGCGGCGGC	CTCTCNGGAA	GGCTGACCGC
	51	ACTGCAGCTT	GCAGAACACAG	GTTATCAGAT	TGCACITTTTC	GATAAAGGCT
25	101	GCCGCGCGGG	CGAACACGCC	GCGCCTATG	TTCGCCGCGC	CATGCTCGCG
	151	CCTGCGCGGG	AAGCGGTGGA	AGCCACGCC	GAGTGGTCA	GGCTGGGACG
	201	CGAGANCATC	CGGCTTTGGC	CGGCGCATCG	ATGCCATCTG	AAAACGCTGT
	251	CCATGATGCA	NGAAACAGCG	AGCGTGATTG	TGTGGCACGG	CGAGGACAAA
	301	CTTTATATCA	ACGAGTCTCT	GCGGCATCTG	AAAGCGCGGC	GCGTAGCGGA
	351	TGACNAATAT	GTCCGTTGGC	GCGCGGACGA	CATGCCCGGA	CGGAGACCGC
30	401	AACTCGCGGG	ACGTTTTC	GACGCGCTCT	ACCTGCCACG	CGAAGGCCAG
	451	CTCGACGGCG	GGCAATATT	GTCTGCACTT	GCCGACGCTT	TGGACGAAC
	501	GAACGTCGCC	TGCCATTTGG	AACACGAATG	TGCCCGCGAA	GACTTGCAG
	551	CCCAATACGA	CTGGCTGATC	GACTGCGCGC	GCTACGGCGC	AAAAACCGCG
35	601	TGGAACCAAT	CCCCGAGNNA	NACCAGCACC	CTGCGCGGCA	TACGCGGCGA
	651	AGTGGCGCGG	GTTTACACAC	CCGAAATCAC	GCTCAACGCG	CCGCTGCGCG
	701	TGCTACACCC	GCGCTATCG	CTNTACATCG	CCCGAAGA	AAACNCNCTC
	751	TTCGTATCAT	GCGGACCCCA	AATCGAAAGC	GAAAGCCAA	CACCTGCCAG
	801	CGTGGCTTCC	GGGCTGGAAC	TCTTATCCGC	ACTCTATGCC	GTCCACCCCG
40	851	CTCTGCGGCA	AGCGGACATC	CTCGAATG	ACACCGGCT	GCGGCCACAG
	901	CTCAATCACC	ACAACCCCGA	AATCGTTAC	AACCGGCCCC	GACGCTGAT
	951	TGAAATCAAC	GGCTTTTCC	GCAACGTTT	CATGATCTCC	CCGCGCGTAA
	1001	CCGCGCGCGC	GCTCAGATTG	GCACTGGCAC	TGTTTGACGG	AAAGGANGCG
	1051	CCCGAACCGC	ATGAGAAAG	CGGTTTGGCG	TATATCCGAA	GACAGATATA
	1101	A				

45 This encodes a protein having amino acid sequence <SEQ ID 814>:

	1	MTRIAILGGGLSGRLTALQLAEQGYQIALFDKSCRRGEHAAAYVAAAMLAPAAEAVEATP
	51	PLSNEFVRHLKRGGVADDXI VWRWADDIAEREPLQGGRFSDGIYLPTEQLDGRQLXSALADALDELNVCHWEHECAFE
50	101	LDGRQLXSALADALDELNVCHWEHECAFE DLQAQYDWLI DCRGYGAKTA
	151	WNQSPKXTSTLRIGRGEVARVYTPETITLNR PVRLHPRYP LYIAPKENKV
	201	LVIGATQIES EQAPASVRS GLELLSALYA VHPAFGEADI LEIATGLRPT
	251	LNHHNPEIRY NRARRLEIN GLFRHGFHMIS PAVTAAAVRL AVALFDGKXA
	301	PERDEESGLA YIRRDQ*
	351	

ORF126a and ORF126-1 show 95.4% identity in 366 aa overlap:

		10	20	30	40	50	60
55	orf126a.pep	MTRIAILGGGLSGRLTALQLAEQGYQIALFDKSCRRGEHAAAYVAAAMLAPAAEAVEATP					
	orf126-1	MTRIAILGGGLSGRLTALQLAEQGYQIALFDKSCRRGEHAAAYVAAAMLAPAAEAVEATP					
		10	20	30	40	50	60
60							

-469-

		70	80	90	100	110	120
	orf126a.pep	EVVRLGRQXIPLWRGIRCHLKTPEAMXKENGSLIVWHGQDKPLSNEFVRHLKRGGVADDXI					
5	orf126-1	EVVRLGRQSIPLWRGIRCRINTHTMMQENGSLIVWHGQDKPLSSEFVRHLKRGGVADDEI					
		70	80	90	100	110	120
		130	140	150	160	170	180
10	orf126a.pep	VRWRADDIAEREPEQLGGRFSDGIYLPTEGQLDGRQILSALADALDELNVPCHEWEHECAPE					
	orf126-1	VRWRADDIAEREPEQLGGRFSDGIYLPTEGQLDGRQILSALADALDELNVPCHEWEHECVPE					
		130	140	150	160	170	180
		190	200	210	220	230	240
15	orf126a.pep	DLQAQYDWLIDCRGYGAKTAWNQSFXXTSTLRGIRGEVARVYTPETILNRPVRLHLHPRYP					
	orf126-1	GLQAQYDWLIDCRGYGAKTAWNQSPEHTSTLRGIRGEVARVYTPETILNRPVRLHLHPRYP					
		190	200	210	220	230	240
20	orf126a.pep	LYIAPKENXVVFVIGATQIESESQAPASVRSGLLELLSALYAVHPAFGEADILEIATGLRPT					
	orf126-1	LYIAPKENXVVFVIGATQIESESQAPASVRSGLLELLSALYAIHPAFGEADILEIATGLRPT					
		250	260	270	280	290	300
25	orf126a.pep	LNHHNPEIRYNRRRLIEINGLFRHGFMSIPAVTAAVRLAVALFDGKXAPERDEESGLA					
	orf126-1	LNHHNPEIRYNRRRLIEINGLFRHGFMSIPAVTAAARLAVALFDGKADAPERDKESGLA					
30		310	320	330	340	350	360
	orf126a.pep	YIRRQDX					
	orf126-1	YIRRQDX					

Homology with a predicted ORF from *N.gonorrhoeae*

ORF126 shows 90% identity over a 180 aa overlap with a predicted ORF (ORF126ng) from *N.gonorrhoeae*:

40	orf126.pep	MTRIAVLGGGLSGRLTALQLAEQGYQIALFDKSCRRGEHAAAYVAAAMLAPAAAXTEATP	60
	orf126ng	MTRIAVLGGGLSGRLTALQLAEQGYQIELEFDKGTGRGEHAAAYVAAAMLAPAAAEVATP	60
	orf126.pep	EVVRLGRQSIPLWRGIRCRINTHTMMQENGSLIVWHGQDKPLSSEFVRHLKRGGXTDDEI	120
45	orf126ng	EVIRLGRQSIPLWRGIRCRINTHTMMQENGSLIVWHGQDKPLSSEFVRHLKRGGVADDEI	120
	orf126.pep	VRWRADDIAEREPEQLGGRFSDGIYLPTEGQLDGRQLSALADALDELNVPCHEWEHECVPE	180
	orf126ng	VRWRADDIAEREPEQLGGRFSDGIYLPTEGQLDGRQILSALADALDELNVPCHEWEHECAFP	180

An ORF126ng nucleotide sequence <SEQ ID 815> was predicted to encode a protein having amino acid sequence <SEQ ID 816>:

```

1  MTRIAVLGGGLSGRLTALQLAEQGYQIELEFDKGTGRGEHAAAYVAAAMLAPAAAEVATP
51  PAEAEVATP EVIRLGRQSI PLWRGIRCRNLNLTMMQENGSLIVWHGQDK
101  PLSSEFVRHLKRGGVADDEI VRWRADDIAEREPEQLGGRFSDGIYLPTEGQ
151  LDGRQLSALADALDELNVPCHEWEHECAFP DLQAQYDQV DCRGYGAKTA
201  WNQSPEHTSTLRGIRGEVARVYTPETILNRPVRLHLHPRYP
251  SSSAREKSKAKAKEPPATVPWNYSFASMPSTPESAKPTS SKWRPGLRET
301  LNHHNPEIRYRERRLIEINGLFRHGFMSIPAVTAAVRLAVALFDGKDA
351  FERDEESGLAYIRQD*
```

Further work revealed the following gonococcal DNA sequence <SEQ ID 817>:

1 ATGACCCGTA TCGCCCTCCT CGGAGGCGGC CTTTCGGGAA GGCTGACCGC
 51 ATGCGAGCTT CGAGAACAAAG GTTATCAGAT TGAACCTTTT GACAAGGGCA
 101 CCGCCCAAGG CGAACACGCC CGCGCTATG TTGCGCGCGC GATGCTCGCG
 151 CTTGCGGCGG AAGCGGTGCA GGCAACGCC GAAGTCATCA GGCTGGGCGAG
 201 GCAGAGCATT CCGCTTTGGC GCGGCATCG ATGCCGCTG AACACGCTCA
 251 CGATGATGCA GGAAACAGGC AGCCTGATTG TGTGGCAGCG GCAGGACAAAG
 301 CCATTATCCA GCGAGTTGCT CCGCATCTC AAACGCGCGG CGGTAGCGGA
 351 TACGGAATC GTCCGTTGGC GCGCGATGA AATCGCGGAA CGCGAACCGC
 401 AACTCGGCGG ACGTTTTTCA GACGCGATCT ACCTCGCGAC CGAAGCGCAG
 451 CTCGAGCGGC GGCAATATTT GTCTGCACTT GCGGACGCTT TGGACGAATC
 501 GAACGTCCTT TGCCATTGGG AACACGAATG GCGCCCCCAA GACCTCGAAG
 551 CCCAATACGA CTGGGTAACT GACTGCGGGG GCTACGCGCG GAAAACCGCG
 601 TGGAAACCAAT CCCCCGAGCA CACGAGCACC TTGCGGGGCA TACGCGGGCA
 651 AGTGGCGCGG GTTTACAGCG CGGAATCAC GCTCAACCGC CCGTSGCGCC
 701 TGTGTCACCC GCGCTATCCG CTCTACATCG CCGCGAAAGA AAACACGCTC
 751 TTGCTCATCG GCGCGACCCA AATCGAAGCG GAAGCGCAG CCCCCGCAAC
 801 CGTAGCTTCC GCGCGTGAAC TCTTATCCG CCGCTATCCG GTCCACCCCG
 851 CTTGCGGCGA AGCGGACATC CTCGAAATCG CGCGCGGCGT GCGCCCCACG
 901 CTCAACGACC ACAACCCCGA AATCCGCTAC AGCGCGGAAC GCGCGCTCAT
 951 CGAAATCAAC GCGCTTTTCC GGACCGGCTT TATGATTTC CCGCGCTAA
 1001 CCGCGCCGCG CGTCAGATTG GCAGTGGCAC TGTTTGACGG AAAAGACGGG
 1051 CCGGACCTG ATGAAGAAAG CGGTTTGGCG TATATCGGAA GACAAGATTA
 1101 A

This corresponds to the amino acid sequence <SEQ ID 818; ORF126ng-1>:

1 MTRIAVLGGG LSGRLTALQL AEQSYQIELF DKGTQRGHEA AAYVAAAMLA
 51 PAEAEVEATP EVIRLGRQSI PLWRGIRCLR NTLTMMQENG SLIVVHGQDK
 101 PLSSEFFVRHL KRGGVADDEI VRWRADEIAE REPQLGGRFS DGIYLPTEGQ
 151 LDGRQILSAL ADALDELNVP CWHWEHCAPQ DLQAGYDWWI DCRGYGAKTA
 201 WNQSPEHTST LRGIRGEVAR VYTPETITLNR FVRLHLPRYP LYIAPKENHV
 251 FVIGATQIES ESQAPASVRS GLELLSALYA VHPAFGEADI LEIAAGLRPT
 301 LNHHPFIRY SRERRLLIEIN GLFRHGFMSI PAVTAAARVL AVALFDGKDA
 351 FERDEESGLA YIGRQD*

ORF126ng-1 and ORF126-1 show 95.1% identity in 366 aa overlap:

		10	20	30	40	50	60
35	orf126-1.pep	MTRIAILGGGLSGRLTALQLAEQSYQIALDKGCRRGHEAAYVAAAMLAPAAEVEATP					
	orf126ng-1	MTRIAVLGGGLSGRLTALQLAEQSYQIELFDTGTRQGEHAAYVAAAMLAPAAEVEATP					
		10	20	30	40	50	60
40	orf126-1.pep	EVVRLGRQSIPLWRGIRCLRNLTHMMQENGSLIVVHGQDKPLSSEFFVRHLKRGGVADDEI					
	orf126ng-1	EVIRLGRQSIPLWRGIRCLRNLTHMMQENGSLIVVHGQDKPLSSEFFVRHLKRGGVADDEI					
		70	80	90	100	110	120
45	orf126-1.pep	VRWRADIAEREPQLGGRFSDGIYLPTEGQLDGRQILSALADALDELNVCHWEHECVPE					
	orf126ng-1	VRWRADEIAEREPQLGGRFSDGIYLPTEGQLDGRQILSALADALDELNVCHWEHECAPQ					
		130	140	150	160	170	180
50	orf126-1.pep	GLQAGYDWWIDCRGYGAKTAWNQSPEHTSTLRGIRGEVARVYTPETITLNRFPVRLHLPRYP					
	orf126ng-1	DLQAGYDWWIDCRGYGAKTAWNQSPEHTSTLRGIRGEVARVYTPETITLNRFPVRLHLPRYP					
		190	200	210	220	230	240
55	orf126-1.pep	LYIAPKENHVFFVIGATQIESESQAPASVRSGLELLSALYAVHHPAFGEADILEIATGLRPT					
	orf126ng-1	LYIAPKENHVFFVIGATQIESESQAPASVRSGLELLSALYAVHHPAFGEADILEIAGLRPT					
		250	260	270	280	290	300
60	orf126-1.pep	LNHHPFIRYNRARRRLIEINGLFRHGFMSIPAVTAAARLVALFDGKDAFERDEESGLA					
	orf126ng-1	LNHHPFIRYNRARRRLIEINGLFRHGFMSIPAVTAAARLVALFDGKDAFERDEESGLA					
		310	320	330	340	350	360
65	orf126-1.pep	LNHHPFIRYNRARRRLIEINGLFRHGFMSIPAVTAAARLVALFDGKDAFERDEESGLA					
	orf126ng-1	LNHHPFIRYNRARRRLIEINGLFRHGFMSIPAVTAAARLVALFDGKDAFERDEESGLA					

orf126ng-1 LNNHNPEIRYSRRRLIEINGLFRHGFMISSPAVTAATAVRLAVALFDGKDAPERDEESGLA
310 320 330 340 350 360

5 orf126-1.pep YIRRQDX
||| ||||
orf126ng-1 YIGRQDX

Furthermore, ORF126ng-1 shows homology to a putative *Rhizobium* oxidase flavoprotein:

gi|2627327 (AF0044408) putative amino acid oxidase flavoprotein [Rhizobium etli]
Length = 327
Score = 169 bits (423), Expect = 3e-41
Identities = 112/329 (34%), Positives = 163/329 (49%), Gaps = 25/329 (7%)

Query: 3 RIAVLGGGLSRLTALQALAEQGYQIELFDKGTROGEHXXXXXXXXXXXXXXXXXXXX 62
RI V G G++G A QL G+++ L ++ G
Sbjct: 2 RILVNGAGVAGLTVAWQLYRHGFVTLAERAGTVGA-GASGFAGGMLAPWCERESAEEFV 60

Query: 63 IRLGRQSIPLWRGIRCLRNLITMMQENGSLIVWHGQDKPLSSEFVRHLKRGGVADIEVR 122
+ LGR + W + G+L+V G+D F R G DE+
Sbjct: 61 LTLRLAAADWEEA-----LPGHVHRRGTLLVAVGGRTGELDRFSRRTS-GWEWLDEVA- 113

Query: 123 WRADIEAEREPIQLGGFRFSDGIYLPTEGQLDGRQLSALADALDELNVPCHEHECAPQDL 182
IA EP L GRF ++ E LD RQ L+ALA L++ + +
Sbjct: 114 -----IAALEPDLAGRFRRALFFRQEAHLDPRLAALAAAGLEDARMRLTLG---VVGES 165

Query: 183 QAQYDWDVICRGYAKTAWNQSPEHTSTLRGIRGEVARVYTPITLNRPVRLLLHPRIPLY 242
+D V+DC G LRG+RGE+ V T E++L+RPVRLLLHPR+P+Y
Sbjct: 166 DVDHDRVVDCTGAA-----QIGRLPLGRVGRGEMLCVETTEVSLSRPVRLLLHPRPIY 218

Query: 243 IAPKENHVFVIGATQIESESQAPASVRSGLLELLSALYAVHFAFGEADIEIAAGLRPTLN 302
I P++ + F++GAT IES+ P + RS +ELL+A YA+HFAFGEA + E AG+RP
Sbjct: 219 IVPDRKNRMFMVAGTAMIESDDGGPIFARSMLLELLNAAAYAMHFAFGEARVTEAGVGRPAYP 278

Query: 303 HNNPEIRYSRRRLIEINGLFRHGFMISSP 331
+ P R ++E R + +NGL+RHGE++P
Sbjct: 279 DNLP--RVTQEGRTLHVNGLYRHGFLLAP 305

This analysis suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 97

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 819>:

1 ATGACTGATA ATCGGGGGTT TACGCTGGTT GAATTAATAT CAGTGGTCTT
51 GATATTGTCT GTACTTGCTT TAATTGTTTA TCCGAGCTAT CGCAATTATG
101 TTGAGAAAGC AAGATAAAT GCAGTGGCGG CAGCCTTGTT AGAAAATGCA
151 CATTTTATGG AAAAGTTTAA TCTGCAGAAAT GGGAGGTTTA AACAAACATC
201 TACCAAGTGG CCAAGTTTCC CGATTAAAGA GGCAGAGGCC TTTTGTATCC
251 GTTTGATAGG AATCGCCGCG CCGG .GCTT TAGACAGTAA ATTCACTGTG
301 AAGCCGCTAG CATACAGATA ACATAAAAAT CCTTTTATTA TTAAGATGAA
351 TGAATAACTTA GTAACCTTTA GTTTGCAAGA AGTCCCGCAG TTCGTGTAGT
401 GACGGCGCTG ATTATTTTAA AGGAAATGAT AAGGACTGCA AGTTACTTAA
451 GTAG

This corresponds to the amino acid sequence <SEQ ID 820; ORF127>:

1 MTDNRGFTLV ELISVVLILS VLALIVYPYS RNYVEKAKIN AVRAALLENA
51 HFMEKFIYQN GRFKQTSTKW PSLPIKEAEG FCIRLINGIVA RXALDSKFML
101 KAVAIKDKNK PFIKMNENL VTFICKKSAS SCSGGLDYFK GNDKDKLLK
151 *

Further work revealed the following DNA sequence <SEQ ID 821>:

```

1   ATGACTGATA ATCGGGGGTT TACGCTGGTT GAATTAATAT CAGTGGTCTT
51  GATATTGTCT GTACTTGCTT TAATTGTTTA TCCGAGCTAT CGCAATTATG
101 TTGAGAAAGC AAGAGTAAAT GCAGTCGGGG CAGCCTTGTT AGAAATATGCA
151 CATTTTATGG AAAAGTTTTA TCTGCAGAAAT GCGAGSTTTA AACAAACATC
201 TACCAAATGG CCAAGTTTTC CGATTAAAGA GGCAGAGGCG TTTTGTATCC
251 GTTTGAATGG AATCGCGCGC GGGGCTTAG ACAGTAATTT CATGTTGAAG
301 GCGGTAGCCA TAGATAAAGA TAAAAATCCT TTTATTATTA AGATGAATGA
351 AAATCTAGTA ACCTTTATTT GCAAGAACTC CGCCAGTTCG TGTAGTGACG
401 GGCTGGATTA TTTTAAAGGA AATGATAAGG ACTGCAAGTT ACTTAAGTAG

```

This corresponds to the amino acid sequence <SEQ ID 822; ORF127-1>:

```

1   MTDNRGFTLV ELISVVLILS VLALIVYPSY RNYVEKAKIN AVRAALLENNA
51  HFMEKFYLQN GRFKQTSTKW PSLPIKEAEG FCIRLNGIAR GALDSKFMFLK
101 AVAIDKDKNP FIIKMNNENLV TFICKKSASS CSDGLDYFKG NDKDKLLK*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF127 shows 98.0% identity over a 150aa overlap with an ORF (ORF127a) from strain A of *N. meningitidis*:

```

20  orf127.pep   MTDNRGFTLVELISVVLILSVLALIVYPSYRNYVEKAKINAVRAALLENNAHFMEKFYLQN
      orf127a    MTDNRGFTLVELISVVLILSVLALIVYPSYRNYVEKAKINAVRAALLENNAHFMEKFYLQN
      10         20         30         40         50         60
25  orf127.pep   GRFKQTSTKWPSLPIKEAEGFCIRLNGIVARXALDSKFMFLKAVAIKDKNPFIIKMNNENL
      orf127a    GRFKQTSTKWPSLPIKEAEGFCIRLNGI-AGGALDSKFMFLKAVAIKDKNPFIIKMNNENL
      70         80         90         100        110        120
30  orf127.pep   VTFICKKSASSCSDGLDYFKGNKDKLLKX
      orf127a    VTFICKKSASSCSDGLDYFKGNKDKLLKX
      130        140        150
35  orf127a      120        130        140        150

```

The complete length ORF127a nucleotide sequence <SEQ ID 823> is:

```

1   ATGACTGATA ATCGGGGGTT TACGCTGGTT GAATTAATAT CAGTGGTCTT
51  GATATTGTCT GTACTTGCTT TAATTGTTTA TCCGAGCTAT CGCAATTATG
101 TTGAGAAAGC AAGAGTAAAT ACAGTCGGGG CAGCCTTGTT AGAAATATGCA
151 CATTTTATGG AAAAGTTTTA TCTGCAGAAAT GCGAGATTTA AACAAACATC
201 TACCAAATGG CCAAGTTTTC CGATTAAAGA GGCAGAGGCG TTTTGTATCC
251 GTTTGAATGG AATCGCGCGC GGGGCTTAG ACAGTAATTT CATGTTGAAG
301 GCGGTAGCCA TAGATAAAGA TAAAAATCCT TTTATTATTA AGATGAATGA
351 AAATCTAGTA ACCTTTATTT GCAAGAACTC CGCCAGTTCG TGTAGTGACG
401 GGCTGGATTA TTTTAAAGGA AATGATAAGG ACTGCAAGTT ACTTAAGTAG

```

This encodes a protein having amino acid sequence <SEQ ID 824>:

```

1   MTDNRGFTLV ELISVVLILS VLALIVYPSY RNYVEKAKIN TVRAALLENNA
51  HFMEKFYLQN GRFKQTSTKW PSLPIKEAEG FCIRLNGIAR GALDSKFMFLK
101 AVAIDKDKNP FIIKMNNENLV TFICKKSASS CSDGLDYFKG NDKDKLLK*

```

ORF127a and ORF127-1 show 99.3% identity in 149 aa overlap:

```

      orf127a.pep   MTDNRGFTLVELISVVLILSVLALIVYPSYRNYVEKAKINAVRAALLENNAHFMEKFYLQN
      orf127-1      MTDNRGFTLVELISVVLILSVLALIVYPSYRNYVEKAKINAVRAALLENNAHFMEKFYLQN

```

		10	20	30	40	50	60
5	orf127a.pep	70	80	90	100	110	120
		GRFKQSTKWPSLP	IKAEAGFCIRL	NGIARGALDSKF	MLKAVAI	DKDNPF	IIKMNNLV
	orf127-1	GRFKQSTKWPSLP	IKAEAGFCIRL	NGIARGALDSKF	MLKAVAI	DKDNPF	IIKMNNLV
		70	80	90	100	110	120
10	orf127a.pep	130	140	150			
		TFICKKSASSCS	SDGLDYFKG	NDKDKCLLKX			
	orf127-1	TFICKKSASSCS	SDGLDYFKG	NDKDKCLLKX			
		130	140	150			

15 Homology with a predicted ORF from *N.gonorrhoeae*

ORF127 shows 97.3% identity over a 150 aa overlap with a predicted ORF (ORF127ng) from

N.gonorrhoeae:

20	orf127.pep	MTDNRGFTLV	ELISVVLIS	VLALIVYPSY	RNYVEKAKIN	AVRAALLEN	AHFMEKFYLQN	60
	orf127ng	MTDNRGFTLV	ELISVVLIS	VLALIVYPSY	RNYVEKAKIN	AVRAALLEN	AHFMEKFYLQN	60
	orf127.pep	GRFKQSTKWPSLP	IKAEAGFCIRL	NGIARXALDSKF	MLKAVAI	DKDNPF	IIKMNNLV	120
	orf127ng	GRFKQSTKWPSLP	IKAEAGFCIRL	NGI-ARGALDSKF	MLKAVAI	DKDNPF	IIKMNNLV	119
25	orf127.pep	VTFICKKSASSCS	SDGLDYFKG	NDKDKCLLKX	150			
	orf127ng	VTFICKKSASSCS	DRLDYFKG	NDKDKCLLKX	149			

The complete length ORF127ng nucleotide sequence <SEQ ID 825> is:

30	1	ATGACTGATA	ATCGGGGGTT	TACACTGGTT	GAATTAAAT	CAGTGGTCTT	
	51	GATATTGTCT	GTACTTGCTT	TAATTGTTT	TCCGAGCTAT	CGCAATTATG	
	101	TTGAGAAAGC	AAAGATAAAT	GCAGTGGGG	CAGCCTTGT	AGAAAAATGCA	
	151	CATTTTATGG	AAAGATTTTA	TCTGCAGAA	GGGAGATTTA	AACAAACATC	
35	201	TACCAAAATG	CAAAGTTTGC	CGATTAAAGA	GGCAGAGGC	TTTTGTATCC	
	251	GTTTGAATGG	AATCGCGCGC	GGGGCTTAG	ACAGTAAATT	CATGTTGAAG	
	301	CGCGTAGCCA	TAGATAAAGA	TAAAAATCCT	TTTATTATTA	AGATGAATGA	
	351	AAATCTAGTA	ACCTTTATTT	GCAAGAACTC	CGCCAGTTCG	TGTAGTGACG	
	401	GGCTGGATTA	TTTTAAAGGA	AATGATAAGG	ACTGCAAGTT	ACTTAAGTAG	

This encodes a protein having amino acid sequence <SEQ ID 826>:

40	1	MTDNRGFTLV	ELISVVLIS	VLALIVYPSY	RNYVEKAKIN	AVRAAFLENA
	51	HFMEKFYLQN	GRFKQSTKW	PSLP	IKAEAG	FCIRLNGIAR
	101	AVAI	DKDNPF	IIKMNNLV	TFICKKSASS	CSDDL
						YDFYFKG
						NKDKCLLKX*

ORF127ng and ORF127-1 show 100.0% identity in 149 aa overlap:

		10	20	30	40	50	60
45	orf127-1.pep	MTDNRGFTLV	ELISVVLIS	VLALIVYPSY	RNYVEKAKIN	AVRAALLEN	AHFMEKFYLQN
	orf127ng-1	MTDNRGFTLV	ELISVVLIS	VLALIVYPSY	RNYVEKAKIN	AVRAALLEN	AHFMEKFYLQN
		10	20	30	40	50	60
50	orf127-1.pep	70	80	90	100	110	120
		GRFKQSTKWPSLP	IKAEAGFCIRL	NGIARGALDSKF	MLKAVAI	DKDNPF	IIKMNNLV
	orf127ng-1	GRFKQSTKWPSLP	IKAEAGFCIRL	NGIARGALDSKF	MLKAVAI	DKDNPF	IIKMNNLV
		70	80	90	100	110	120
55		130	140	150			
	orf127-1.pep	TFICKKSASSCS	SDGLDYFKG	NDKDKCLLKX			
	orf127ng-1	TFICKKSASSCS	SDGLDYFKG	NDKDKCLLKX			
		130	140	150			
60		130	140	150			

This analysis, including the fact that the predicted transmembrane domain is shared by the meningococcal and gonococcal proteins, suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

5 Example 98

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 827>

```

1      .GTGTCGCTGG  CTTGCGTGAT  TGCCTCTCAA  ATCTTCCTTT  ACGAAGATT
51     CAACCAAAATG  CGGAAAACCG  GTGGAGCTAT  CTGCGGTTTT  CTTGTCCAA
101    ATTTATCTCGG  GGTTCAGCA  GGGGATTATC  GATTTGAGTG  CCGACGAGAA
151    CCCCGTACTG  CATATCTGGT  CTTTGGCAGT  AGAGGAACAG  TATTACTCTC
201    TGTATCCCTT  TTTGCTGATA  TTTTGTCTGA  AAAAAACCAA  ATCGCTACGG
251    GTGCTGCGTA  ACATCAGCAT  CATCTCGATT  TTGATTTTGA  CTGCCTCATC
301    GTTTTGTCCA  AGCGGCTTTT  ATACCGGATT  CTTCAACCAA  OCCATACTTT
351    ATTACCTCTT  GACACTGAGG  TTTCCGAGC  TGTGTCAGG  TTGCTGCTGT
401    CGGGTTTACG  GGCNAACGCA  AAACGCGACA  CGGCAACAG  CAATGGGAAA
451    ACGGCACTTG  CTTTCTATCAC  TCTGCTTCGG  CGCATTGCTT  GCCTGCTGTG
501    TCGTGATTGA  CAACACAAAT  CCGTTATCCC  CGGGAATGAC  CCGTCTCTCT
551    CCTGCTCTGC  TGACGGCACT  GCTTATCCGG  AGTATGCAAT  ACGGGACACT
601    TCCGACCCGC  ATCCTGTGG  CAGCCCCCAT  GTATTGTGTC  GGCAAAATCT
651    CTTATTCCCT  ATACTGTAC  CATTGGATT  TTATTGCTTT  CGCTCCGCTC
701    ATTAGAGGCG  GGAAACAGCT  CGGACTGCCT  GCCG..

```

This corresponds to the amino acid sequence <SEQ ID 828; ORF128>:

```

1      .VSLASVIASQ  IFLYEDFNQM  RKTVELSAVF  LSNLYLGFQQ  GYFDLSADEN
51     PVLHIWSLAV  EEQYLLLYPL  LLIFCCCKTK  SLRVLRNISI  ILFLILTFSS
101    FLPSGFYTDI  LNQNTYYLS  TLRFPPELLAG  LLRLAVYQQT  NGRRQTANGK
151    RQLLSSLCFG  ALLACLFLVID  KHNFPFIPMT  LLLPCLLTAL  LIRSMQYGTL
201    PTRILSASPI  VEVGKISYSL  YLYHWFTAF  APLIRGGKQL  GLPA..

```

Further work revealed the complete nucleotide sequence <SEQ ID 829>:

```

1      ATGCRAAGCTG  TCCGATACAG  ACCGGAATTT  GACGGATTGC  GGGCGCTGCG
51     CGTGCTATTC  GTCATGATTT  TCCACTGATA  TAACCGCTAG  CTGCGCGRAG
101    GATTCTCTGG  GGTGACATTT  TTCTTGTGCA  TCTCAGGATT  CCTCATTACC
151    GGCAATCAITC  TTCTGAAAT  ACAGAACGGT  TCTTTTCTT  TCCGGGATTT
201    TTATACCCGC  AGGATTAAGC  GGATTATATC  TGCCTTTATT  GCGGCGGTGT
251    CGCTGGCTTC  GGTGATTGCC  TCTCAAATCT  TCTTTTAGCA  AGATTTCAC
301    CAATGCGGGA  AAACCGTGGA  GCTTTCTGCG  GTTTTCTGT  CCAATATTTA
351    TCTGGGGTTT  CAGCAGGGGT  ATTTCGATT  GAGTGGCGAC  GAGAACCCCG
401    TACTGCATAT  CTGCTCTTGT  GCAGTAGAGG  AACAGTATTA  CCTCCTGTAT
451    CCCCTTTTGC  TGATATTTTG  CTGCAAAAAA  ACCAAATCGC  TACGGGTGCT
501    GCGTAACATC  AGCATCATCC  TGTITTTGAT  TTTGACTGCC  TCATCGTTT
551    TGCCAAAGCG  GTTTTATACC  GACATCTCTA  ACCAACCCTA  TACTTATTAC
601    CTTTCGACAC  TGAGGTTTCC  CGAGCTGTG  GCAGGTTTCG  TGCTGGCGGT
651    TTRGGGCGAA  ACGCAAAACG  CGAGACGGCA  AACAGCAAT  AGAAACCGCG
701    AGTTGCTTTC  ATCACTCTGC  TTGCGCGCAT  TCTTTCGCTG  CTGTCTCGTG
751    ATTGACAAAC  ACATCTCTT  TATCCCGGGA  ATGACCTCTG  TCGCTCTGCG
801    CTTCTGACG  GCACCTGTTA  CTGCGAGTAT  GCATACGGG  ACATCTCCGA
851    CCGCATCTCT  GTGCGCAAGC  CCGATCTGAT  TTGTGCGGCA  AATCTCTTAT
901    TCCCTATACC  GTTACCATTG  GATTTTTATT  GCTTTCGCC  ATTACATTAC
951    AGGCGACAAA  CAGCTCGGAG  TGCTGCGCT  ATCGGCGGTT  GCCGCGTTGA
1001  CGGCGCGATT  TTCCCTGTGG  AGTTATTATT  TGATTGAACA  GCGCTTTAGA
1051  AAACGGAAGA  TGACCTTCAA  AAAGGCATTT  TCTGCTCTCT  ATCTCGAAC
1101  GTCCCTGATA  CTTGTCGGTT  ACAACCTGTA  CGCAAGGGGG  ATATTGCAAC
1151  AGGAACACCT  CCGCCCGTTG  CCGGCGCGCG  CCCTTGCTGC  GGAAATTCAT
1201  TTTCCGGAAC  CGCTCTGAC  CTGCGGAGC  TCGCACGCGC  GACACTGAG
1251  GGGGTTTCTG  GATTATGTGC  GCAGCCGGGA  AGGTTGGAAA  GCACAAATCC
1301  TGTCCCTCGA  TTCGAGGTGT  TTGGTTTGG  TAGATGAGAA  GCTGCGCAGC
1351  AACCCGTAT  GTCGAAAAAT  CCGGATGAA  GTTGAAAAAG  CCGAAGCCGT

```

1401 TTTCATTGCC CAATTCTATG ATTGAGGAT GGGCGGCCAG CCTGTGCCGA
 1451 GATTTGAAGC GCAATCCCTC CTAATACCCG GGTTCOCAGC CCGATCAGG
 1501 GAAACCGCTCA AAGGATAGC GCCTCTCAA CCGCTCTATG TTTTTCGAAA
 1551 CAACACATCA ATCAGCGGTT GCCCCTCAG GAGGAGAAAA TTGAAAAGAT
 1601 TTGCGCGAAA CCAATATCTC GCGCCCATTC AGGCTATGGG CGACATCGGC
 1651 AAGAGCAATC AGGCGGTCTT TGATTTGATT AAGATATTC CCAATGTGCA
 1701 TTGGTGGGAC GCACAAAAAT ACCTGCCCAA AAACACGGTC GAAATATACG
 1751 GCCGCTATCT TTACGGCGAC CAAGACCACC TGACCTATTT CGGTTCCTAT
 1801 TATATGGGGG GGAATATCCA CAAACACGAA CGCTCGCTTA AATCTTCCCA
 1851 CGGCGGGGCA TTGCAGTAG

This corresponds to the amino acid sequence <SEQ ID 830; ORF128-1>:

1 MQAVRYRPEI DGLRAVAVLS VMIFHLNNRW LPGAFLGVDI FVISGFLIT
 51 GIILSEIQNG SFSFRDFYTR RIKRIYPAFI AAVSLASVIA SQIFLYEDFN
 101 QMRKTVLSA VFLSNYILGF QQGYFDLSAD ENPVLHIWSL AVESQYLLY
 151 PLLLIFFCKK TKSLRVLNRI SIILFLILTA SSFLPSGFYT DILNQNTYV
 201 LSTLRFPPELL AGSLLAVYQG TQNGRRCTAN GRQQLSSSLC FGALLACLVE
 251 IDKHNFPIFG MTLLEPLLIL ALLIRSMOYG TLPFRLLSAS FIVFVGKTSY
 301 SLYLHWIIFI AFPHYTGDKK QLGFLPVSAY AALTAGFSLL SYLIEQPLR
 351 KRKMTKEAF FCYLYLAPSL LVGYNLYARG TLKOEHLREL PGAPLAENH
 401 FFEVLTGLD SHAGHLRGFL DYVGSRECKW AKILSLDSEC LVNWDEKLD
 451 NPLCRKYRDE VEKAEAVFIA QFYDLRMGGQ PVPRFEAQSF LIPGFPARF
 501 ETVKRIAIAVK PVYVFANNTS ITRSPLRCEK LKRFAANQYL RPIQAMGDIG
 551 KSNQAVFDLI KDIPNVHVD AQKYLFPKNTV EYGRYLYGD QDHLTYFGSY
 601 YMGREFHKE RLLKSSHGGA LQ*

Computer analysis of this amino acid sequence gave the following results:

Homology with hypothetical integral membrane protein HI0392 of *H. influenzae* (accession number U32723)

ORF128 and HI0392 show 52% aa identity in 180aa overlap:

Orf128: 1 VSLASVIA SQIFLYEDFNQMRKTVLSAVFLSNYILGFQQGYFDLSADENPVLHIWSLAV 60
 +L S IAS IP+Y DEN++RKT+EL+ FLSN YLG QGYFDLSA+ENPVLHIWSLAV
 HI0392: 46 MALVSFIASAIIFYNDENKLRKTIELATAFLSNFYLGTLTQGYFDLSADENPVLHIWSLAV 105
 Orf128: 61 EEQXXXXXXXXXIFCCKKTKSLRVLNRSIILFLILTASSFLPSGFYTILNQNPNTYYLS 120
 E Q I KK + ++VL I ++LFL IL A+SF+ + FY ++L+QPN YYLS
 HI0392: 106 EGQYILIFELLILAYKKFREVKVLFITLILFEILLATSFVSANFYKEVLHQPNIYYLS 165
 Orf128: 121 TLRFPPELLAGSLAVYQGQTQNGRRQTANGKRLQLSSLCFGALLACLVEIDKHNPFPFGMT 180
 LRPELL GSLLA+Y N + Q + +L+ L L +CLF++ + FPG+T
 HI0392: 166 NLRFPPELLVGSLLAIYHNLSN-KVQLSKQVNNILAILSTLLFLSCLFIMNNNIATIPGIT 224

Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF128 shows 98.0% identity over a 244aa overlap with an ORF (ORF128a) from strain A of *N.*

meningitidis:

orf128.pep 10 20 30 VSLASVIA SQIFLYEDFNQMRKTVLSAVF
 orf128a ILSEIQNGSFSFRDFYTRIKRIYPAFIAAVSLASVIA SQIFLYEDFNQMRKTVLSAVF
 60 70 80 90 100 110
 orf128.pep 40 50 60 70 80 90
 LSNIYILGFQQGYFDLSADENPVLHIWSLAVEEQYLLYPLLLIFCCKKTKSLRVLNRSI
 orf128a LSNIYILGFQQGYFDLSADENPVLHIWSLAVEEQYLLYPLLLIFCCKKTKSLRVLNRSI
 120 130 140 150 160 170
 orf128.pep 100 110 120 130 140 150
 ILFLILTASSFLPSGFYTILNQNPNTYYLSTLRFPPELLAGSLAVYQGQTQNGRRQTANGK
 orf128a ILFLILTATSFPSGFYTILNQNPNTYYLSTLRFPPELLAGSLAVYQGQTQNGRRQTANGK
 180 190 200 210 220 230

-476-

		160	170	180	190	200	210
	orf128.pep	RQLSSSLCFGALLACLEFVIDKHNFFIPGMTLLLPCLLTALLIRSMQYGLTLPTRILSASPI					
5	orf128a	RQLSSSLCFGALLACLEFVIDKHNFFIPGMTLLLPCLLTALLIRSMQYGLTLPTRILSASPI					
		240	250	260	270	280	290
	orf128.pep	VFVGKISYSLYLYHWIFIAFAHIFAPLIRGGKQLGLPA					
10	orf128a	VFVGKISYSLYLYHWIFIAFAHYITGDKQLGLPAVSAVALTAGFSLLSYYLIEQLRKR					
		300	310	320	330	340	350
	orf128a	KMTFKFAFFCLYLAPSLILVGYNLYARGILKQEHRLRPLPGAPLAAENHFFETVLTGLDSSH					
		360	370	380	390	400	410

15 The complete length ORF128a nucleotide sequence <SEQ ID 831> is:

	1	ATGCAAGCTG	TCCGATACAG	ACCGGAATTT	GACGGATTGC	GGGCCGTGCG
	51	CGTGCATCC	GTCCATGATT	TCCACTGGA	TAACCGCTGG	CTGCCGCGAG
	101	GATTCCTGGG	GGTGGACATT	TCTCTGTCA	TCTCGAGATT	CTCATTAC
	151	GGCATCATTC	TTTCTGAAT	ACAGAAGCTG	TCTTTTCTT	TCCGGGATT
20	201	TTATACCCGC	AGGATTAAGC	GGATTATCC	TGCTTTTAT	CGCGCGGTG
	251	CGCTGGCTTC	GGTGATTGCC	TCTCAAATCT	TCCTTTACGA	AGATTTCAC
	301	CAAAATGCGGA	AAACCGTGG	GCTTCTGCG	GTTTCTTGT	CCAATTATTA
	351	TCTGGGGTTT	CAGCAGGGGT	ATTTGCGATT	GAGTCCGCG	GAGAACCCCG
25	401	TACTGCATAT	CTGGCTTTTG	CGAGTAGAG	AACAGATTA	CCTCGTGAT
	451	CCTCTTTTGC	TGATATTTTG	CTGCAAAAA	ACAAAATCGC	TACGGGTGCT
	501	GGGTAAACATC	AGCATCATCC	TATTCTGTAT	TTGACTGCC	ACATCGTTTT
	551	TGCCAAGCGG	GTTTATATAC	GATATTCTCA	ACCAACCCAA	TACTTATAC
	601	CTTTCGACAC	TGAGGTTTCC	CGAGCTGTG	CGAGGTTCCG	TGCTGGCGGT
30	651	TTACGGGCAA	ACGCAAAACG	GCAGACGGCA	AACAGCAAT	GGAAAACGGC
	701	AGTTGCTTTC	ATCACTCTGC	TTCCGGCGAT	TGCTTGCGT	CGTGTGTG
	751	ATTGACAAAC	ACATCTGTT	TATCCCGGGA	ATGACCGCTG	TCTTCCCTG
	801	CTCGCTGAGC	CGACTCTCTA	TCCGAGTAT	GCATATCGG	ACACTTCGGA
	851	CCCGCATCCT	GTCCGCAAGC	CCCATGATAT	TTGTCGCAA	AATCTCTAT
35	901	TCCCTATACC	TGTACCATG	GATTTTATT	GCTTTCGCG	ATTACATTAC
	951	AGGCGACAAA	CAGCTCGGAC	TGCTGCGGT	ATCGGCGGT	GCGCGTTGA
	1001	CGGCGGATT	TTCCCTGTG	AGTTATTATT	TGATTGAACA	GCCGCTTGA
	1051	AAACCGGAAGA	TGACCTTCAA	AAAGGCATT	TTCTGCCTCT	ATCTCGCCCG
	1101	GTCCCTGATA	CTTGTGGTT	ACAACCTGTA	CGCAAGGGGG	ATATTGAAC
40	1151	AGGAACACCT	CCGCGCGTTG	CCCGGCGCG	CCCTTGCTGC	GGAAAATCAT
	1201	TTTCCGAAAA	CCGTCTGAC	CTCGGCGAC	TGCGACCGCG	GACACCTCG
	1251	GGGGTTCTG	GATTATGTCG	GCAGCGGGGA	AGGGTGAAAA	GCCAAATTC
	1301	TGTCCTCGA	TTCCGAGTGT	TTGGTTTGG	TAGATGAGAA	CTGGCGACG
	1351	ACCTCTTAT	GTGCAAAATA	CCGGGATGAA	GTTGAAAAAG	CCGAGCGCGT
	1401	TTTCATCTGC	CAATCTATG	ATTGAGAGAT	GGCGCGCAG	CCGTCGCGA
45	1451	GATTTGAGC	GCAATCCTTC	CTATACCGG	GCTTCGACG	CCGATCTAG
	1501	GAACCGTCA	AAAGGATAGC	CGCGTCAAA	CCCGCTATG	TTTTGCAAA
	1551	CAACACATCA	ATCAGCCGTT	CGCCCTGAG	GGAGGAAAA	TTGAAAAGAT
	1601	TTGCGCGAAA	CCAATATCTC	CGCCCATCTG	AGGCTATGGG	GCACATCGC
50	1651	AAGAGCAATC	AGGCGGTCTT	TGATTTGATT	AAAGATATAT	CCAATGTGCA
	1701	TTGGGTGGAC	GCACAAAAAT	ACCTGCCCAA	AAACACGGTC	GAAATATAGC
	1751	CGCGCTATCT	TTACGGCGAC	CAAGACCACC	TGACCTATTT	CGGTCTCTAT
	1801	TATATGGGGC	GGGAATTCCA	CAACACGAA	CGCTGCTTA	AATCTTCTCG
	1851	CGACGGCGCA	TTGCAGTAG			

This encodes a protein having amino acid sequence <SEQ ID 832>:

55	1	MQAVRYRPEI	DGLRAVAVLS	VMIFHNNRW	LPGGFLGVDI	FFVISGFLIT
	51	GIILSEIQNG	SFSERDPYTR	RIKRIYPAFI	AAVSLASVIA	SOIPLYEDFN
	101	QMRKTVELSA	VELSNVITAG	QQGYFELSD	ENFVLHIWSL	AVEEQYLYLY
	151	PLLIFCCKK	TKSLRVLRLN	SEILFLITGA	TSGLSPGYIT	DILNQNTYIT
60	201	LSLRFPELL	AGSLIAYVQG	TQNGRRQTAN	GRQLLSLIC	TGALLACLFV
	251	IDKHNPFPIPG	MTLLLECLLT	ALLIRSMQYG	TLTRILSAS	PIYFVGKISY
	301	SLYLYHWIFI	AFAHYITGDK	QLGLPAVSA	AVLTAGFSL	SYLYLIEQLR
	351	KRKMTFKFAF	FCLYLAPSLI	LVGYNLYARG	ILKQEHRLR	PGAPLAAENH
	401	FPETVLTGSD	SHAGHLRGFL	DYVSGREGWK	AKILSLDSEC	LWVWDEKID
	451	NPLCRKYRDE	VEKAEAVFIA	QFYDLRMGGQ	EVPRFEAQSF	LIPGFAPRAR
65	501	ETVKRIAIAVK	PVTVFANNST	ISRSPLREEK	LKRFAANQYL	RPIQAMGDIG
	551	KSNQAVFDLI	KDIPNVHWVD	AQRKYLPRKNT	EIYGRYLVGD	QDHLTYFGSY

601 YMGRFHKHE RLLKSSRDGA LQ*

ORF128a and ORF128-1 show 99.5% identity in 622 aa overlap:

	orf128a.pep	MQAVRYRPEIDGLRAVAVLSVMIFHLNNRWLPGGFLGVDIFFVISGFLITGIILSEIQNG
5	orf128-1	MQAVRYRPEIDGLRAVAVLSVMIFHLNNRWLPGGFLGVDIFFVISGFLITGIILSEIQNG
	orf128a.pep	SFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTVELSAVFLSNIIYLGF
	orf128-1	SFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTVELSAVFLSNIIYLGF
10	orf128a.pep	QQGYFDLSADENPVLHIWSLAVEEQYYLLYPLLLIFCCKKTKSLRVLRNISIIILFILTA
	orf128-1	QQGYFDLSADENPVLHIWSLAVEEQYYLLYPLLLIFCCKKTKSLRVLRNISIIILFILTA
15	orf128a.pep	TSFLPSGFYTDILNQNTYYLSTLRFPPELLAGSLAVYQGTQNGRRQTANGKROLLSSLC
	orf128-1	TSFLPSGFYTDILNQNTYYLSTLRFPPELLAGSLAVYQGTQNGRRQTANGKROLLSSLC
20	orf128a.pep	FGALLACL FVIDKHNPPIPGMTLLPCLLTALLIRSMQYGTLPTRILSASPIVEVGKISY
	orf128-1	FGALLACL FVIDKHNPPIPGMTLLPCLLTALLIRSMQYGTLPTRILSASPIVEVGKISY
	orf128a.pep	SLYLHWIFIAFAHYITGDKQLGPAVSAVAALTAGFSLLSYLIEQPLRKRKMTFKKAF
25	orf128-1	SLYLHWIFIAFAHYITGDKQLGPAVSAVAALTAGFSLLSYLIEQPLRKRKMTFKKAF
	orf128a.pep	FCLYLAPSLILVGYNLYARGILKQSHLRPLPGAPLAAENHPETVITLGDSSHAGHLRGFL
	orf128-1	FCLYLAPSLILVGYNLYARGILKQSHLRPLPGAPLAAENHPETVITLGDSSHAGHLRGFL
30	orf128a.pep	DYVGSREGWKAKILSLDSECLVWVDEKLADNPLCRKYRDEVEKAEAVFTAQFYDLRMGGQ
	orf128-1	DYVGSREGWKAKILSLDSECLVWVDEKLADNPLCRKYRDEVEKAEAVFTAQFYDLRMGGQ
35	orf128a.pep	PVPRFEAQSLIPGFPARFRET VKRIA AVKPVYVFANNTSISRSPLEEEKLKRFAANQYL
	orf128-1	PVPRFEAQSLIPGFPARFRET VKRIA AVKPVYVFANNTSISRSPLEEEKLKRFAANQYL
40	orf128a.pep	RPIQAMGDIGKSNQAVFDLIKIDPNVHWVDAQKYLPKNTVEIYGRYLGDDHLYTFGSY
	orf128-1	RPIQAMGDIGKSNQAVFDLIKIDPNVHWVDAQKYLPKNTVEIYGRYLGDDHLYTFGSY
	orf128a.pep	YMGRFHKHERLLKSSRDGALQX
45	orf128-1	YMGRFHKHERLLKSSHGAGLQX

Homology with a predicted ORF from *N.gonorrhoeae*

ORF128 shows 93.4% identity over 244 aa overlap with a predicted ORF (ORF128ng) from *N. gonorrhoeae*:

50	orf128.pep	VSLASVIASQIFLYEDFNQMRKTVELSAVF	30
	orf128ng	ILSEIQNGSFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTIELSTVF	112
55	orf128.pep	LSNIIYLGFQQGYFDLSADENPVLHIWSLAVEEQYYLLYPLLLIFCCKKTKSLRVLRNISI	90
	orf128ng	LSNIIYLGFRLGYFDLSADENPVLHIWSLAVEEQYYLLYPLLLIFCYKTKTKSLRVLRNISI	172
	orf128.pep	ILFILITASSFLPSGFYTDILNQNTYYLSTLRFPPELLAGSLAVYQGTQNGRRQTANGK	150
60	orf128ng	ILFILITASSFLPAGFYTDILNQNTYYLSTLRFPPELLVGSLLAVYQGTQNGRRQTENGK	232
	orf128.pep	RQLSSSLCFCGALLACL FVIDKHNPPIPGMTLLPCLLTALLIRSMQYGTLPTRILSASPI	210
	orf128ng	RQLSSSLCFCGALLVCL FVIDKHDPFIPGTLTLLPCLLTALLIRSMQYGTLPTRILSASPI	292

orf128.pep VFGKISYSLYLYHWIFIAFAPLIRGGKQLGLPA 244
||||| | | | | | | | | | | | | | | | | | | |
orf128na VFGKISYSLYLYHWIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLLSYYLIEQPLRKR 352

5 The complete length ORF128ng nucleotide sequence <SEQ ID 833> is:

1	ATGCAAGCTG	TCGGATACAG	GCGTGA AATT	GACGGATTGC	GGGCGCTGCG
51	CGTGCTGTCG	GTGATATT	TCACCATGAA	TACGCTGTCG	CTGGCCGGAG
101	GATGTCCTGG	GCTGGACATT	TCCTTTGTCA	TCCTGGGATT	CCCTCATTACC
151	AATCATCACT	TTGTGAAAT	CACAGACGCG	CTCTTTTCTT	TCGGGGATTG
201	TACATGACGC	GGTGAATACG	GGAATTATCC	TCGCTTTATT	GGCGCGCTGG
251	CCCTGGCTTC	GGTGAATGCT	TCCTAAATCT	TGCTTTACGA	AGATTTCACAC
301	CAAAATGAGA	AAACCTATAG	GCTTTCTACG	TTGTTTTTGT	CCAAATATTAT
351	TTTGGGTTCT	GCTTGGGGTG	ATTTCGATT	GAGTGCGCAG	AGAAACCCCG
401	TACTGTCAT	TCGATTTCTG	GGCGGTAGAG	ACGAGTTAT	CCCTCGTATG
451	CTCTTCTTTC	TGCTGTTCTG	TTCGAAAAAA	ACCAATTAC	TACGGGTGCT
501	GGCTAATATC	AGCATATCCC	TCCTTCTGAT	TTTGACGCCA	CTATCTGTTT
551	TGCGCGCGCG	GTTTTATATC	GACATCTCTA	ACCAACCCaa	TACTATTATC
601	CTCTGCGAC	TGAGGTTTCC	CGAGCTGTGG	GTCGGGTTGC	TGTTGGCGGT
651	TATGCGCGCA	AGCAAAACCG	CGACAGCGTC	ACGAGAAAT	GGAAAACGCG
701	AGTGTCTTTC	ATTACTCTTG	TTCCGGCGCT	tgctTGTCTG	CTGTGTCGTG
751	ATGCACAAC	ACGATCCGTT	TATCCCGGGA	ATAACCCCTG	TCCTCTCTTC
801	CTGCTGACG	GGGCTCGTTA	TCCGGAGTAT	CGCAATACGG	ACACTGCTGC
851	CCCGATCTCT	TCTGGCAGAT	TCCTGCGGTC	TTCTTGAGCA	ATATCTTAT
901	TCCTCTATAC	CGCTGCTGAC	GATTTTATT	GCGTCCGCC	ATATCATATC
951	AGCGGCAAT	CAGCTCGGAC	GCGCTGCCGT	ATCGCGCGCT	CGCGGCTAGA
1001	CGGCGCGATT	TTCCCTCTTG	AGCTATTTT	TGATTGAAGCA	CGCGCTTGA
1051	AAACCGGAAG	TGACCTTCAA	AAGGCAATT	TTCCTGCTTT	ATCTCGGCC
1101	GTCCCTGATG	CTCTGCTGGT	ACAACCTGTA	TTCAGAGGG	GATTGTAAC
1151	AGGAACACCT	CGCGCGCGTG	CCGCGACGCG	CGGTTGCTGC	GGAAAATATG
1201	TTTCCGGAAA	CCGCTCTGAC	CTCTGCGCAC	TGCGACGCG	CGACCTTCGC
1251	GGGGTTCGTG	GATTATTGCG	CGCGCAGGGA	AGGGTGGAAA	GCTAAATCTG
1301	TGTCCTCTCG	TCGCGGATGT	TGCGTGTGG	AGGATGAGAA	CTGCGACAG
1351	AACCCGTTGT	CGCGAAAAAT	CTGGGATGAA	TGTAAAAAG	CCGAAGCTGC
1401	TTTCACTGCC	CATTCTATG	ATTGAGGAT	GGGCGGCCAG	GGGCGCTGGA
1451	GTTTGTGAAG	CGAATCCCTC	CTGATACGAG	GGTTCAAGG	CGGATCTGAG
1501	GAAACCGCTA	AGAGATATCT	CGCGCTGAT	CTTGGAGCA	TTTGGAGTGA
1551	CATTACCTAT	CGCTGCGTT	CTCTCTGAG	GAGCAAGCA	TTTGAAGAAT
1601	TGCTGTCAAA	CGATACCTCT	GCGGCTATTC	GGGCTATGCG	CGACATCGCG
1651	AGAGACAACT	AGGCGGTCCT	TGATTGTGTT	AAAGATATGG	CAATGTGTGA
1701	TTGGTGGGAC	GCACAAAAAT	ACCTGCCCAA	AAACACGGTC	GAATATACAG
1751	GACGCTATCT	TTACGGCGAC	AGGCCAACCC	TGACCTATT	CGGTTCTTCT
1801	TATGTTGGGC	GGGAATTCTA	CACACACGAA	CGCCTGCTCA	AGCATTCCCG
1851	AGGCGGCGCA	TGTCAGTAG			

This encodes a protein having amino acid sequence <SEQ ID 834>:

45	1	MOAVRYRPEI	DGSLRAVAVLS	VILPHLNNK	LPGGGLGVDI	FFVISGFLTI
	51	NLLSYRQNG	SGSRFRDYFR	RIKRIYPAFI	AAVSLASVIA	SOFLFYEDFN
	101	QMRKTIELST	VFLSNLYLRF	RLGIFYDLSD	ENFVHLTWSL	AVEEQYXLYL
	151	PLLILFCYKK	TKSLRNLVGI	ILSLFILFTA	SSFLPAGVFT	DINNOPTNY
	201	LSTLRPFPELL	VGSLIYAVGQ	TONGRQTRN	GKROLLSLIC	FGALLVCLVF
50	251	IDKHDFPPIFG	ITLLPCLLIT	ALLRSOMQV	PTILTRILAS	PVIFVGRISY
	301	SLYLIWHYFI	APAFYITGDK	QGLGPAVAVS	AALTATGAPL	SYLLIEQPIR
	351	KRKMTFKKAF	KLVLAPLMS	LVGNLYISRG	IKQIHSRL	PCFPAVNN
	401	FFEVLTGLD	SVAGHRLPL	AKILDSIC	WVDEKDLA	ILPGKFKARF
	451	NPLCRVATFI	DEKRAVFTA	QYDLRMGQ	PVPEREAQSL	ILPGKFKARF
55	501	ETVTKRIAIVK	PHYVFNANTS	ISRSPLREK	LKRFANILY	RPRAMGDPI
	551	KSNQAVFDLV	KDIPNVHWD	AQKYLPKNTV	EIRGRYLYGD	QDHLTYFGSY
	601	YNGEFPHKHE	RLKHSRSGA	LQ*		

ORF128ng and ORF128-1 show 95.7% identity in 622 aa overlap:

60 orf128-1.pep MQAVRYRPEIDGLRAVAVLSVMIFHLNNRWLPGGLGVDDIFFVISGFLITGIILSEIQNG
| | | | |
orf128ng MQAVRYRPEIDGLRAVAVLSVIFHLNNRWLPGGLGVDDIFFVISGFLITNIILSEIQNG

orf128-1.pep SFSFRDFYTRRKIKRIYPAFIAAVSLASVASQIFLYEDENQMRKTVELSAVFLSNYLGF


```

orfl28ng      |||||
SFSFRDFTYTRRIKRIYAFIAAASLASVIASQIFLYEDFNQMRKTIELSTVFLSNIYLG
5  orfl28-1.pep QQGYFDLSADENPVLHIWLSLAVEEQYYLLYPLLLIFCCCKTKSLRVLRNISIIFLILTA
orfl28ng      |||||
RLGYFDLSADENPVLHIWLSLAVEEQYYLLYPLLLIFCYKTKSLRVLRNISIIFLILTA
orfl28-1.pep SSFLPSGFTDILNQPNYYLSTLRFPPELLAGSLLAVYGGTQNGRRQTANGKRQLSSLC
10 orfl28ng      |||||
SSFLPAGFTDILNQPNYYLSTLRFPPELLVGSLLAVYGGTQNGRRQTENGKRQLSSLC
orfl28-1.pep FGALLACLFLVIDKHNPFIPGMTLLLPCLLTALLIRSMQYGTLPTRILSASPIVFVGKISY
15 orfl28ng      |||||
FGALLVCLFLVIDKHDPFIPGITLLLPCLLTALLIRSMQYGTLPTRILSASPIVFVGKISY
orfl28-1.pep SLYLHWIFIAFAHYITGDKQLGLPAVSAAVATAGFSLLSYLIEQLPRKKRMTFKKAF
orfl28ng      |||||
SLYLHWIFIAFAHYITGDKQLGLPAVSAAVATAGFSLLSYLIEQLPRKKRMTFKKAF
20 orfl28-1.pep FCLYLAPSLILVGNLYARGILKQEHRLPLPGAPLAAENHFPETVLTLDGSHAGHLRGFL
orfl28ng      |||||
FCLYLAPSLMLVGNLYSRGILKQEHRLPLPGTVAEAENHFPETVLTLDGSHAGHLRGFL
orfl28-1.pep DYVGSREGWKAKILSDSECLVWVDEKLADNPLCRKYRDEVEKAEAVFIAQFYDLRMGGQ
25 orfl28ng      |||||
DYVGSREGWKAKILSDSECLVWVDEKLADNPLCRKYRDEVEKAEAVFIAQFYDLRMGGQ
orfl28-1.pep PVPRFEAQSLIPGFPARFRETVKRIAARKVYVYFANNTSISRSLPREEKLRKFAANQYL
30 orfl28ng      |||||
PVPRFEAQSLIPGFKARFRETVKRIAARKVYVYFANNTSISRSLPREEKLRKFAANQYL
orfl28-1.pep RPTQAMGDIKGSNQAVFDLHKDIPNVHWDQAQYLPKNTVEIHGRYLYGDQDHLTYFGSY
35 orfl28ng      |||||
RPIRAMGDIKGSNQAVFDLVKIDIPNVHWDQAQYLPKNTVEIHGRYLYGDQDHLTYFGSY
orfl28-1.pep YMGREFHKHERLLKSSHGGALQX
orfl28ng      |||||
YMGREFHKHERLLKHSRGGALQX
610 620

```

In addition, ORF218ng shows homology to a hypothetical *H. influenzae* protein:

```

sp|P43993|Y392_HAEIN  HYPOTHETICAL PROTEIN HI0392 >gi|1074385|pir||B64007
hypothetical protein HI0392 - Haemophilus influenzae (strain Rd KW20)
>gi|1573364 (U32723) H. influenzae predicted coding region HI0392 [Haemophilus
influenzae] Length = 245
45 Score = 239 bits (604), Expect = 3e-62
Identities = 124/225 (55%), Positives = 152/225 (67%), Gaps = 1/225 (0%)

Query: 38 VDIFEVISGFLITNIIILSEIQNGSFSFRDFTYTRRIKRIYPXKXXXXXXXXXXXXXXFLYE 97
+DIFEVISGFLIT II++EQ SFS + FYTRRIKRIYP F+Y
50 Sbjct: 1 MDIFEVISGFLITGIIITEIQNSFSLSKQFYTRRIKRIYPAFITVMALVSFIASAIFTYN 60

Query: 98 DFNQMRKTIELSTVFLSNLYLGLRGLYFDLSADENPVLHIWLSLAVEEQXXXXXXXXXIFC 157
DEN++RRTIEL+ FLSN YLG GYFDLSA+ENPVLHIWLSLAVE Q I
55 Sbjct: 61 DFNKLRKTIELATAIATLSNLYLGLTQGYFDLSANENPVLHIWLSLAVEGQYYLIFFLILILA 120

Query: 158 YKKTSLRVLRNISIIILFLILTASSFLPAGFTDILNQPNYYLSTLRFPPELLVGSLLAV 217
YKK ++VL I++IL IL A+SF+ A FY ++L+QPN YLS LRFPELLVGSLLA+
Sbjct: 121 YKFFREVKVLFIITLILFILLATSFSVSANFYKEVLHQPNYYLSNLRFPPELLVGSLLAI 180

60 Query: 218 YGQTQNGRRQTENGKRQLLSLFCGALLVCLFLVIDKHDPFIPGIT 262
Y N + Q +L++L L CLF+++ + FPGIT
Sbjct: 181 YHNLSN-KVQLSKQVNNILAILSTLLFSCFLMMNNNIAFIPGIT 224

```

This analysis, including the identification of several putative transmembrane domains, suggests that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 99

- 5 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 835>:

```

1  ..ATTATTACG AATACCGCTG GATGTTTCTT TACGGCGCAC TGACGACCTT
51  GGGGCTGACG GTCGTGGCAA C.GGGGGCGG TTCGTATTG GTCCTGTGTT
101  TGGCGTTGGC GCGCCTGATT CACTTGGAAA AAGCCGGTGC GCCGATGGCG
151  CTGCTGGCGT GGGCGTTGCG TAAAGTTTCG CTGCTGTATG TTACGCTGTT
201  CCGGGGTACG CCGCTGTTTG TGCAGATTGT GATTGGGGCG TATGTGTGTT
251  TTCCGTTTTT CCTC..

```

This corresponds to the amino acid sequence <SEQ ID 836; ORF129>:

```

1  ..IIYEYRMWFL YGALTTLGLT VVAXAGGSVL GLLLLALARLI HLEKAGAPMR
51  VLAWALRKVS LLYVTLFRGT PLFVQIVWA YVWFFFEV..

```

- 15 Further work revealed the complete nucleotide sequence <SEQ ID 837>:

```

1  ATGGATTTC GTTTGACAT TATTACGAA TACCGCTGA TGTTCCTTA
51  CGGCGCACTG ACGACCTTGG GGCTGACGGT CGTGGCAACG GCGGGCGGTT
101  CGGTATTGGG TCTGTTGTTG GCGTTGGCGC GCCTGATTCA CTTGAAAAAA
151  GCGCGTGGCG CGATGCGCGT GCTGGCGTGG GCGTGGCTGA AAGTTTCGCT
201  GCTGTATGTT ACGCTGTTCC GGGGTACGCC GCTGTTTGTG CAGATTGTGA
251  TTGGGGCGTA TGTGTGTTT CCGTTTTCG TCGATCCTTC AGACGGCATT
301  TTGCTCAGCG GCGAGGCGCG AATCGCGCTG CTGCGCGGAT ACGGGCGGCT
351  GATTGCGGCT TCTTTGGCAC TGATCGCCAA CTCGGGGGCG TATATCTGTG
401  AGATTTTCCG CAGTCTGATG ACGAAGGACA GATGAGGCG
451  GCGCGTTCTT TGGGCTGAC CTATCGCGAG CGATCGCGCT ATCGATTCT
501  GCGCGAGCCA TTGCGGCCCA TCCTGCCCGC TTGCGGAGC GAGTTCATCA
551  CGCTCTTGAA AGACAGCTCG CTGCTGTCGG TCATTGCTGT GCGGAGATTG
601  GCGTATGTTT AGAATACGAT TACGGGCGCG TATTGCGTTT ATGAGAGACC
651  GCTTTACACC GTGCGCCCTGA TTTATCTGTT GATGACGACT TTCTTAGGCT
701  GGATATTCTT GCGTTTGAA AAACGTTACA ATCGGCAACA CCGCTGA

```

This corresponds to the amino acid sequence <SEQ ID 838; ORF129-1>:

```

1  MDFRFDIIYE YRMWFLYGAL TTLGLTVVAT AGGSVLGLLL ALARLIHLEK
51  AGAPMRVLAW ALRKVSLLYV TLFRTPLFV QIVWAYVWF PFFVHPSDGI
101  LVSGEAAIAL RRGYGLIAG SLALIANSGA YICEIFRAGI QSIDKGQMEA
151  ARSLGLTYPO AMRYVILPQA LRRLMLPLAS EFTTLIKDSS LLSVIAVAEL
201  AYWQNTITGR YSVYEEPLYT VALIYLLMTT FLGWIFLRLE KRYNPQHR*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF129 shows 98.9% identity over a 88aa overlap with an ORF (ORF129a) from strain A of *N.meningitidis*:

```

10      20      30      40      50
orf129.pep  IIYEYRMWFLYGALTTLGLTVVAXAGGSVLGLLLLALARLIHLEKAGAPMRVLAW
45          |||||
orf129a     MDFRFDIIYEYRMWFLYGALTTLGLTVVATAGGSVLGLLLLALARLIHLEKAGAPMRVLAW
          10      20      30      40      50      60
60      70      80
orf129.pep  ALRKVSLLYVTLFRGTPLEFVQIVWAYVWFFFEV
          |||||

```

orf129a ALRKVSLLYVTLFRGTPLEFVQIVIWAYVWFFVFHPSDGLVSGEAAIALRRGYGPIIAG
 70 80 90 100 110 120

5 orf129a SLALIANSGAYICEIFRAGIQSIDKGQMEAAARSLGLTYPQAMRYVILPQALRRMLPPLAS
 130 140 150 160 170 180

The complete length ORF129a nucleotide sequence <SEQ ID 839> is:

1 ATGGATTTC GTTTTGACAT TATTACGAA TACCGCTGA TGTTCCTTA
 51 CGGCGCACTG ACGACCTTGG GCGTGACCGT CGTGGCGACG CGCGGCGGTT
 101 CGGTATTGGG TCTGTTGTTG CGGTGGCGCG GCGTGATTCA CTGGAAAAA
 151 GCGGTCGCGC CGATGCGCGT GCTGGCGTGG GCGTGCCTGA AGGTTTCGCT
 201 GCTGTATGTT ACGCTGTTCC GGGGTACGCC GCTGTTTG GAGATTGTA
 251 TTGGGGCGTA TGTGTGTTT CCGTTTTCG TCCATCCTTC AGACGGCATT
 301 TTGTTAGCG CCGAGGCGCG AATCGCGCTG CCGCGCGGAT ACGGGCGCT
 351 GATGCGCGCT TCTTTGGCAC TGATCGCCAA CTCGGGCGCG TATATCTGTG
 401 AGATTTTCCG CCGGCGCATC CAGTCTATAG ACBAAGGACA GATGCGAGCG
 451 GCGGCTTCTT TGGGCTGAC CTATCCGACG CGATCGCGCT ATGCTATTCT
 501 GCGCGAGGCA TTGCGCGCTA TGCTGCGCCG TTTGGCGAGC GAGTTCATCA
 551 CGCTCTTGAA AGACAGCTCG CTGCTGTGCG TCATTGCTGT GCGCGAGTTG
 601 GCGTATGTTC AGAATACGAT TACGGGCGCG TATTGCGTTT ATGAGAACC
 651 GCTTTACACC GTCGCCCTGA TTTATCTGTT GATGACGACT TTCTTAGGCT
 701 GGATATTCTT GCGTTTGAA AACGTTACA ATCCGCAACA CCGCTGA

This encodes a protein having amino acid sequence <SEQ ID 840>:

1 MDRFRDIIYE YRWMFLYGAL TTLGLTVVAT AGGSLVGLLL ALARLIHLEK
 51 AGAEMRVLAW ALRKVSLLYV TFRGTPLEFV QIVIWAYVWF PFFVHPSDGI
 101 LVSGEAAIAL RRGYGPIIAG SLALIANSGA YICEIFRAGI QSIDKGQMEA
 151 ARSLGLTYPQ AMRYVILPQA LRRMLPPLAS EFTLLKDS LSVIAVAEFL
 201 AYVQNTITGR YSVYEEPLYT VALIYLMTT FLGWIFLRL KRYNPQHR*

ORF129a and ORF129-1 show 100.0% identity in 248 aa overlap:

orf129a.pep MDRFRDIIYEYRWMFLYGALTTLGLTVVATAGGSLVGLLLALARLIHLEKAGAPMRVLAW
 30 orf129-1 MDRFRDIIYEYRWMFLYGALTTLGLTVVATAGGSLVGLLLALARLIHLEKAGAPMRVLAW
 orf129a.pep ALRKVSLLYVTLFRGTPLEFVQIVIWAYVWFFVFHPSDGLVSGEAAIALRRGYGPIIAG
 35 orf129-1 ALRKVSLLYVTLFRGTPLEFVQIVIWAYVWFFVFHPSDGLVSGEAAIALRRGYGPIIAG
 orf129a.pep SLALIANSGAYICEIFRAGIQSIDKGQMEAAARSLGLTYPQAMRYVILPQALRRMLPPLAS
 40 orf129-1 SLALIANSGAYICEIFRAGIQSIDKGQMEAAARSLGLTYPQAMRYVILPQALRRMLPPLAS
 orf129a.pep EFTLLKDSLSLVIAVAELAYVQNTITGRYSVYEEPLYTVALIYLMTTFLGWIFLRL
 45 orf129-1 EFTLLKDSLSLVIAVAELAYVQNTITGRYSVYEEPLYTVALIYLMTTFLGWIFLRL
 orf129a.pep KRYNPQHRX
 orf129-1 KRYNPQHRX

Homology with a predicted ORF from *N.gonorrhoeae*

ORF129 shows 98.9% identity over a 88 aa overlap with a predicted ORF (ORF129ng) from *N.gonorrhoeae*:

orf129.pep IIEYRWMFLYGALTTLGLTVVAXAGGSLVGLLLALARLIHLEKAGAPMRVLAW 54
 55 orf129ng MDRFRDIIYEYRWMFLYGALTTLGLTVVATAGGSLVGLLLALARLIHLEKAGAPMRVLAW 60
 orf129.pep ALRKVSLLYVTLFRGTPLEFVQIVIWAYVWFFVFV 88
 orf129ng ALRKVSLLYVTLFRGTPLEFVQIVIWAYVWFFVILHTAFLGNAMQSRVVPDKGRWIAG 120

An ORF129ng nucleotide sequence <SEQ ID 841> was predicted to encode a protein having amino acid sequence <SEQ ID 842>:

```

1  MDRFRDIIYE YRWMFLYGAL TTLGLTVVAT AGGSVLGLLL ALARLIHLEK
51 AGAPMRVLAW ALRKVSLLYV TLFRTPLFV QIVWAYVWF PFFVILHTAF
101 LGNAMRQSRSS VPDKGRWAG SLELNCQPRG RKTRGEFFPG ESNLGTEPRN
151 PLSMGGRFP GCENWYPPQN FIKK*

```

Further work revealed the following gonococcal sequence <SEQ ID 843>:

```

1  ATGGATTTTc gTTTTGACAT TATTACgaa TACCGCTGGA TGTTCCTTTA
51 CGGCGCACTG Acgaccttgg ggtgacggt cgtggcgacg gCGGGCGGTT
101 CGGtattggg TCTGTTGTTG CGGTGGCGCG GCCTGATTCA CTGGAAAAA
151 GCCGGTGCSC CGATGCGCGT GCTGGCGTGG GCCTTGCCTA AGGTTTCGCT
201 GCTGTACGTT ACCTGTTTCC GGGTACGCC GCTGTTGTG CAGATTGTGA
251 TTTGGCGGTA TCTGTGGTTT CCGTTTTCG TCCATCCTTC AGAGGCGAT
301 TTGCTCAGCG GCGAGGCGGC AATCGCGCTG CGTCCGGAT AGCGGCCGT
351 GATTGCCGCT TCTTTGGCAC TGATGCCAA CTCGGGGCGC TATATCTGTG
401 AGATTTTCCG CGCGGGCATC CAGTCTATAG ACAAGGACA GATGGAGCG
451 CGGTGTTCTT TGGGACTGAC CTATCCGCG CCGATGCGCT ATGTGATCT
501 GCCGCAGGCA TTGGCGCGTA TGCTGCCGCC TTTGGCGAGC GAGTTCATCA
551 CGCTCTTGAA AGACAGCTCG CTGCTGTCGG TCATTGCTGT GCGCGAGTTG
601 GCGTATGTTC AGAATACGAT TACGGCGCG TATTCTGTT ATGAAGACCC
651 GCTTTACACC GCGGCCCTGA TTTATCTGTT GATGACGACT TTCTTAGGCT
701 GGATATTCTT GCGTTTGAA AAACGTTACA ATCCGCAACA CCGCTGA

```

This corresponds to the amino acid sequence <SEQ ID 844; ORF129ng-1>:

```

1  MDRFRDIIYE YRWMFLYGAL TTLGLTVVAT AGGSVLGLLL ALARLIHLEK
51 AGAPMRVLAW ALRKVSLLYV TLFRTPLFV QIVWAYVWF PFFVHPSDGI
101 LVSGEAAIAL RRGVPLIAG SLALIANSGA YICEIFRAGI QSIDKQMEA
151 ARSLGLTYPO AMRVYLTPQ ARRMVPLAS EFTLLKDS LSVIAEAL
201 AYVQNTITGR YSVYEEPLYT VALIYLLMTT FLGWIFLRL KRYNPQHR*

```

ORF129ng-1 and ORF129-1 show 99.2% identity in 248 aa overlap:

```

30  orf129-1.pep MDRFRDIIYERWMFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf129ng-1 MDRFRDIIYERWMFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||

35  orf129-1.pep ALRKVSLLYVTFLFRGTPLFVQIVWAYVWFFVFHPSDGLVSGEAAIALRRGYGLPIAG
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf129ng-1 ALRKVSLLYVTFLFRGTPLFVQIVWAYVWFFVFHPSDGLVSGEAAIALRRGYGLPIAG
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||

40  orf129-1.pep SLALIANSGAYICEIFRAGIQSIDKQMEAAARSLGLTYPQAMRVYILPQALRRMLPPLAS
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf129ng-1 SLALIANSGAYICEIFRAGIQSIDKQMEAAARSLGLTYPQAMRVYILPQALRRMLPPLAS
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||

45  orf129-1.pep EFTLLKDSLLSVIAEALAYVQNTITGRYSVYEEPLYTVALIYLLMTTFLGWIFLRL
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf129ng-1 EFTLLKDSLLSVIAEALAYVQNTITGRYSVYEEPLYTVALIYLLMTTFLGWIFLRL
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||

orf129-1.pep KRYNPQHRX
      |||||||||
orf129ng-1 KRYNPQHRX

```

In addition, ORF129ng-1 is homologous to an ABC transporter from *A.fulgidus*:

```

50  2650409(AE001090) glutamine ABC transporter, permease protein (glnP)
      [Archaeoglobus fulgidus]length = 224
      Score = 132 bits (329), Expect = 2e-30
      Identities = 86/178 (48%), Positives = 103/178 (57%), Gaps = 18/178 (10%)

```

```

55  Query: 65 VSLLYVTFLFRGTPLFVQIVWAYVWFFVFHPSDGLVSGEAAIALRRGYGLPIAGSLAL 124
      +S YV + RGTPV VQI+I +F P+ GI + E A G +AL
      Sbjct: 58 ISTAYVEVIRGTPLLWQILI-----VYGLPAIGINLQPEPA-----GIIAL 99

```

Query: 125 IANSYAYICEIFRAGIQSIDKGQMEAAACSLGLTYFQAMRYVILFQALRRMLFPLASEFIT 184
 SGAYT EI RAGI+SI GQMEA SLG+TY QAMRYVI FQA R +LF L +EPI
 Sbjct: 100 SICSGAYTAETVRAGIESIPIGQMEAAARSLGMTYLOQAMRYVIFPQARNILPALGNEFIA 159

5 Query: 185 LLKDSLLSVIAVAELAYQVNTITGRYSVYEEPLYTAALYLLMTTFLGWIFLRLEKR 242
 LLKDSLLSVI++ EL V I P AL YL+MT L + +K+
 Sbjct: 160 LLKDSLLSVISIVELTRVGRQIVNTTFNAWTFPLGVALFYLMMTIPLSRSLVAYSQKK 217

This analysis, including the identification of transmembrane domains in the two proteins, suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 100

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 845>:

```

1 ..CTGAAAGAAT GCGCTCTGAA AGACCCTGTT TTTATTCCAA ATATCGTTTA
51 TAAGAACATC GCATTTACTT TCCTGCTCTT GCAGCGCGCC GCGGAACCTT
101 GCGTGCCTCG GCAACCGCGC GGTTTTACCG CGCTCGCGCT CGGCTTCATC
151 CTGCTCGCCA ACCTCGCTGA gCTTCACAT CACGAACCTC TACCTAAACA
201 CTAGCTCGCG ACTTATTACT TGCCTCAACT CTTTGCGCGC GCGAGCTAG
251 TTGTSGACAG GCGCGCGGwA ATTACAAAC CTGCCCGCYT CGCGCGCTCT
301 GCACCTGATT ACCCTCGCGC GCATGATGGG CGGCGTGATG ATGGTCTGGC
351 TGACCGCGCG ACTGTGGCAC AGCGGCTTTA CCAAATCGA CTACGCCAAA
401 CTCTGCGCGA TTGCGGTGCC CATCCTTTTC GCGCGCGCGC TCTCGCGCGC
451 TTTCTTG+TG AACGTGAAC CG+TATTTT CATTAACGTT CCGCGGATTC
501 TGACCGCGCG CGTATTGCTA CTGTATCTTT Tc+GTTTAT ACCGATATTT
551 CGGCGGAATG CGTTTACAGA CGATCCGGA TAr

```

This corresponds to the amino acid sequence <SEQ ID 846; ORF130>:

```

1 ..LKECRKLPDV FIPNIVYKNI AITFLLHAA AELMLPQTA GFTALAVGFI
51 LLAKLRELHH HELLRKHYVR TYLLQLFAA AGSLWTGAAX LQNLPAASAPL
101 HLITLGGMMG GMMVWLTAG LWSHGFTKLD YPKLCRIAPV ILFAAAVSRa
151 FLXNVNEXFF ITVFALLTAA VFVLYLFXFI PTFRANAFTD DPE*

```

Further work revealed the complete nucleotide sequence <SEQ ID 847>:

```

1 ATGCGGCGGT TTTTCGTGG GCGGCGGTGT CTGGCATAC TCGGTGGCGT
51 GGTGTTTTTC ATCAACCCCG GTGCCATCGT CCGTGCACGC CAAATTTTCT
101 TGGAACCTAT GCTGCGCGCG GCATACGCGC GTTTTTGAC TCGGCTTTTG
151 TTGACTGGA CGGGTTTTTC GGGTACCTGT AAACCTGTGC CGACTTTGAT
201 GCGGCGATTA TTGCTCGCGC CATCCGCTAT ACTGCCCTTT TCGCGCGAAA
251 CTGCCCTGTT TTTTCGTGCC GCCTATTGGC TGGTGTGTCT GCTGTTCTGC
301 GCGCGGCTGA TTGGCTAGA CCGAAACACC GACAACCTCG CCCTGCTAAT
351 GTTACTTGCC GCGTTCAGT TTTTTCAGAC GGCATATGCC GTCAGCGCGC
401 ATTTGAACCT GTTGCAGCGC CAAGTGCATC TAATATATGG GCGGTGATG
451 TTGCTATCGG TGGCGGCTCAG TATTCCTTTG GCGCGCGAAG CCCTGAAAGA
501 ATGCGCTCTG AAAGACCGCT TTTTATTCC AAATATCGTT TATAAAAACA
551 TCGCCTATAC TTTCTGTGTC TTGCAGCGCG CCGCGCAACT TTGGCTGGCC
601 GCGCAACCGC CGGTTTTCAC GCGCTCTGCA CTGCGCTTCA TCTGTCTGCG
651 CAGCTCGGCT GAGCTTCAAC ATCAGCAACT CTACCTGATA CACTAGTCTC
701 GCACCTTATA CCGTCTCAA CTCTTTGCGC CCGCAGGCTA TTTCTGGACA
751 GCGCGCGGGA AATTACAAAA CCGCGCGCC CTGCGSCCCC TCGCACTGAT
801 TACCCTCGGC GGCATGATGG GCGGCGTGAT GATGGTGTGG CTGACCGCGC
851 GACTGTGGCA CAGCGGCTTT ACCAAACTCG ACTACCCCAA ACTCTGCGCG
901 ATTGCCGTCC CCATCTCTTT CGCGCGCGCC GTCTGCGCGC CTTTCTTGAT
951 GAACGTGAAC CCGATATTTT TCATTACCGT TCGTGCAGAT TGACCGCGCG
1001 CCGTATTCGT ACTGTATCTT TTCACGTTTA TACCGATATT TCGGGCGAAT
1051 GCGTTTACAG ACGATCCGGA ATAA

```

This corresponds to the amino acid sequence <SEQ ID 848; ORF130-1>:

```

1 MRPFVFGAAV LAILGALVFF INPGAIVLHR QIFLEMLPLA AYGGFLTAAL
51 LDWTGFSGNL KPVATLMAAL LLASATLFF SPQTASFVFA AYWLVLVLLFC

```

101 ARLLWLDNRN DNFALMLLLA AFVVFQYATA VSGDLNLLRA QVHINLMAAMV
151 VSVSVSVSILL GAFAAKZKRL KDPTVPTNIV YKNIATITFLL LHAAGELWLT
201 AQTAGCTATA VGPIILLAKLR ELHHHEHGLR HYRVTYVLLQ LFAAAGVYLTW
251 GAQKQLNLEA SAPLHLITFL GMHGGVVMNV LTAGLWHSGF TKLDYPKLCR
301 IAPVLPFAAA VSRAFLMNVN PIFFITVPAT LFAAVFVLYL FTFTIPIFRN
351 AFTDDEP*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF130 shows 94.3% identity over a 193aa overlap with an ORF (ORF130a) from strain A of *N.*

10 *meningitidis:*

orf130.pep
LKECR LKDPV FIPNIVYKN IAITFLLHAA
| | | | |
orf130a
LNLLRAQVHLNMAAVMFVSVRVSITLLGASALKECRLKDPVFIPNIVYKNIAITFLLHAA
140 150 160 170 180 190

orf130.pep
AELWLPAQTAGFTALAVGVFTILAKLRLEHHHELLRKHYVVTYYLLQLFAAGSLWTGAAX
| | | | |
orf130a
AELWLPAQTAGFTSLAVGVFTILAKLRLEHHHELLRKHYVVTYYLLQLFAAGYLWTGAAX
200 210 220 230 240 250

orf130.pep
LQNLPASAPLHLITLGGMGSMVMWVLTAGLWNHSQTFKLDYPKLCRIARVPILFAAASVRA
| | | | |
orf130a
LQNLPASAPLHLITLGGMGSMVMWVLTAGLWNHSQTFKLDYPKLCRIARVPILFAAASVRA
260 270 280 290 300 310

orf130.pep
FLXNVNXPFFITVPAILTAARVLYLFEXFPIFRANAFTDDPEX
| | | | |
orf130a
VLNVNVPFFITVPAILTAARVLYLFVFIPIRANAFDDPEX
320 330 340 350

The complete length ORF130a nucleotide sequence <SEO ID 849> is:

35	1	ATGCGGCGCTT	TTTTCAGTCGG	CGGCGCGGGT	CTTGCCATAC	TCGGTCGCGCT
	51	GTCGTTTTCCT	ATGTCAGCCGG	CGGCAACCGG	CGGCAACCGG	CAAACTTTGAT
	101	TGGAACTATG	CGTCCGCGCG	GCGATACCGG	GTTTTTTGAC	TGCGGCTGCT
	151	TGGACTGTGA	CGGGTTTTCG	GGGTAATCGT	AAACCTGCGT	CGACTTTGAT
40	201	GGGCGGATTA	TGTCTCGCGG	CACTCCGATAT	AGACTCCCTTT	TCGCGCGCAA
	251	CTGCTTCGTT	TGTTCGCGCC	GCGTATTTGCG	TGGTGTGTGT	GCTTTCCTGG
	301	CGCCGCGCTGA	TTTGGCTAGA	CGGAACACCG	GACAACCTGC	CCCTGCTAAT
	351	GTTACTTGGC	GCGTTCACTG	TTTTCAGAC	GCGCATTTGCC	GTCGCGGGTG
	401	ATTTGAACCT	GTTCGCGCGG	CAAGTGTGATC	TAATATGATG	GCGGCTGATG
	451	TTCGCTATAG	TGCGGCTGAG	TATTCCTTTT	CGGCGCGAAG	CTCGGAAAGA
45	501	ATGCGGCTCT	ATGCGGCTCT	CAATCTGCTG	CAATCTGCTG	TATGAAACAA
	551	TGCGCATATG	CTTCTGCTCT	GTGCAACCGC	TTGCGGCTCT	TCGCGGCTCT
	601	GCGCAAACTC	CGGGTTTATC	TCGCGTCGCC	CTGCGCTTTA	CTGCGCTTGC
	651	CAAGCTTCGT	GAGCTTCAC	CTCAAGCACT	CTCGCGGAAA	CACCTACGTC
	701	GCACTTATTA	CTCTGCTCAA	CTCTTTGCGC	CGCGAGGATA	TTTGTGGACA
50	751	GCGCGCGCGA	GAATACAAAG	CTCGCGCGCC	TCGCGCGGCG	TGCACCTGAT
	801	TACCTTCGGT	ACCTGATGAG	CGACGCTGAT	GATGGTGTCG	CTGACTGCGC
	851	CAGCTGTGCA	CAGCGCGGTT	ACGAAGCTCG	ATCTACCGGA	ACTCTGCGCG
	901	ATCGCGCTCT	CGACGCTTTT	CGCGCGCGCC	GTTTCGCGAA	CTGTTTTPAT
	951	GAACTGCTAT	TGACGACGCT	TGACGACGCT	CCCGCGAAAT	CTGACGCGCG
55	1001	CCGCTGTCCT	CTCTTACCTG	CTGACATCTT	TGCGGCTCTT	TGCGGCGAAC
	1051	GCGTTTACGA	ACGATCCGGA	ATGA		

This encodes a protein having amino acid sequence <SEO ID 850>:

1 MRPFFVGA AV LAILGALVFF INPGAIVLHR QIFLELMLPA AYGGFLTAAL
51 LDWTGFGSGL KPVATLMAAL LLAASAILPF SPOTASEEVA AYWLVLLEF

5 101 ARLIWLDRNT DNFALLMLLA AFTVFQATAYA VSGDLNLLRA QVHNLMAAVM
 151 FVSVRVSILL GAELKECRLL KDPVFIPNVV YKNIATFLL LHAALWLP
 201 AGTAGTSLA VGFTLLAKLR ELHHHELLRK HYVATYLLQ LRAAGYLWT
 251 GAAKLQNLFA SRPLHLITLG GMMGSMVMW LTAGLWHSF TKLDYPKLCR
 301 IAVPILFAAA VSRVLMNVN PIFFTVPAI LTAAVFVLYL LTFVPIFRAN
 351 APTDDPE*

ORF130a and ORF130-1 show 98.3% identity in 357 aa overlap:

10 orf130a.pep MRFPFVGAAVLAILGALVFFINPGAIVLHRQIFLEMLPAAYGGLTAALLDWTGFSGNL
 orf130-1 MRFPFVGAAVLAILGALVFFINPGAIVLHRQIFLEMLPAAYGGLTAALLDWTGFSGNL
 15 orf130a.pep KP VATLMAALLLAASAILPFSPTASFFVAAYWLVL LFCARLIWLDRNTDNFALLMLLA
 orf130-1 KP VATLMAALLLAASAILPFSPTASFFVAAYWLVL LFCARLIWLDRNTDNFALLMLLA
 20 orf130a.pep AFTVFQATAYAVSGDLNLLRAQVHNLMAAVMFVSVRVSILLGAELKECRLLKDPVFIPNVV
 orf130-1 AFTVFQATAYAVSGDLNLLRAQVHNLMAAVMFVSVRVSILLGAELKECRLLKDPVFIPNVV
 25 orf130a.pep YKNIATFLLHAAELWLPAQTAGFTSLAVGFTLLAKLREHLHHELLRKHYVRYTYLLQ
 orf130-1 YKNIATFLLHAAELWLPAQTAGFTSLAVGFTLLAKLREHLHHELLRKHYVRYTYLLQ
 30 orf130a.pep LFAAAGYLWTGAAKLQNLPASAPLHLITLGGMMGSMVMWLTAGLWHSF TKLDYPKLCR
 orf130-1 LFAAAGYLWTGAAKLQNLPASAPLHLITLGGMMGSMVMWLTAGLWHSF TKLDYPKLCR
 orf130a.pep IAVPILFAAAVSRVLMNVNPIFFITVPAILTAAVFVLYL LTFVPIFRANAFTDDPE
 orf130-1 IAVPILFAAAVSRVLMNVNPIFFITVPAILTAAVFVLYL LTFVPIFRANAFTDDPE

Homology with a predicted ORF from *N.gonorrhoeae*

ORF130 shows 91.7% identity over a 193 aa overlap with a predicted ORF (ORF130ng) from *N.gonorrhoeae*:

35 orf130.pep LKECRLLKDPVFIPNVVYKNIATFLLHAA 30
 orf130ng LNLRAQVHNLMAAVMFVSVRVSILLGTETLKECRLLKDPVFIPNVVYKNIAT-LLHAA 201
 40 orf130.pep AELWLPAQTAGFTALAVGFILLAKLREHLHHELLRKHYVRYTYLLQLFAAAGSLWTGAAX 90
 orf130ng AELWLPAQTAGFTALAVGFILLAKLREHLHHELLRKHYVRYTYLLQLFAAAGYLWTGAAX 261
 orf130.pep LQNLPASAPLHLITLGGMMGSMVMWLTAGLWHSF TKLDYPKLCRIAVPILFAAAVSRA 150
 45 orf130ng LQNLPASAPLHLITLGGMTGGVMWLTAGLWHSF TKLDYPKLCRIAVSILFASAVSRA 321
 orf130.pep FLNMNVEXFFITVPAILTAAVFVLYLXFIPIFRANAFTDDPE 193
 orf130ng VLMNVNPIFFITVPEILTAAVFVLYLTFVPIFRANAFTDDPE 364

50 An ORF130ng nucleotide sequence <SEQ ID 851> was predicted to encode a protein having amino acid sequence <SEQ ID 852>:

55 1 MNKFFTHMR PFFVGAAVLA ILGALVFHQ PRRYHPAPPN FLGTYAAGCI
 51 RRFFDYRFVG PDGFFRQPT CRYFDGGVVA CCGCFIAVFT ATCRIFRRL
 101 LAGVAARLRL ADLARQHRT LRSVDVTAAF TVQQTAYAVS GDLNLLRAQV
 151 HLNMAAVFV SVRSVLLGT ETLKECRLLK DPVFIPNVYK NIATFLLHAA
 201 AELWLPAQT AGTALAVGF ILLAKLREHL HHELLRKHYV RTYLLQLQFA
 251 AAGYLWTGAA KLQNLPASAP LHLITLGGMT GGVMWLTAT GLWHSF TKL
 301 DYPKLCRIAV SILFASAVSR AVLNMNVNPI FTFVPEILTA AVFVLYL LTF
 351 VPIFRANAFT DDPE*

Further work revealed the following gonococcal DNA sequence <SEQ ID 853>:

```

1 ATGCGCCCGT TTTTCGTGCG TCGGCCAGTA CTGCGCATAC TCGGTGCGTT
51 GGTGTTTTTT ATCAACCCCG GCGCTATCAT CCTGCACCGC CAAATTTTCT
101 TGGAACTTAT GCTGCCGCGT GCATACGGCG GTTTTTTGAC TACCGCTTTG
151 TTGGACCGGA CGGGTTTTTC AGGCAACCTG AAACCTGCGC CTACTTTGAT
201 GCGCGTGTG TTGCTTGTG CGGCTGTTT ATTGCGGTTT TTACCGCAAC
251 TTGCGCGCAT TTTCGTGCGC GCGCTATTGC TGGTGTTCGT GCTGTTCTGC
301 GCGTGCCTGA TTTGGCTCGA CGGCACACAC GACACTTCG CTCGTGTTGAT
351 GTTACTGCGC CATTTCAGAC TTTTTCAGAC GTTCAGCGCGC GCTCAGCGCGC
101 ATTTGAACCT ACTGCGCGCG CAACTGCATT TGAATTATCG GCGGTCATG
451 TTGCTATCCG TCCGCGCTCAG CGTCTTTTTC GGCACGGAAA CCCTGAAGAA
501 ATGCGGCTCG AAAGACCCCG TATTCATCCC CAACGTTATC TATAAAARCA
551 TCGCCATCAC CCTGCTGCTG CACGCCGCGC CGGAACCTTG GCTGCCCGCG
601 CAAACCGCGC GTTTTACTGC GCTTGCGGTC GCGTTCATCC TGCTCGCCAA
651 GCTGCGCGAA CTGCACCATC ACGAATCTTT ACGCAAAAC TACGTCGGCA
701 CTTATTACCT GCTCCAGCTC TTTGCGCGCG CAGGTATCT GTGCACAGCG
751 GCGGCGAACC TGCAAAACCT GCCGCGCTCC GCGCCCTGCG ACCTGATTAC
801 CCTCGCGCGC ATGACGGGTG GCGTGATGAT GGTGTGGCTG ACTGCGCGAC
851 TGTGGCAGAC CGGCTTTTAC AAACCTGACT ACCCGAACTC CTGCGCGCAT
901 GCGGCTCCA TCCTTTTTCG CTCGCGCGTG TCGCGCGCTG TTTTAATGAA
951 CGTGAATCCG ATATTCTTCA TCACGCTTCC CGAGATTCTG ACCGCGACCG
1001 TTTCATGCTT TTACCTGCTG ACGTTGCTAC CGATTTTTCG AGCGAAGCGC
1051 TTTACAGACG ATCGGAATA A

```

This corresponds to the amino acid sequence <SEQ ID 854; ORF130ng-1>:

```

1 MRPFFVGAVAL LAILGALVFF INPGAII LHR QIFLEMLPA AYGGLTTLAL
51 LDRTFGSGNL KPAATLMAVL LLVAALLPFE LPQLAAFFVA AYLVLVLLCG
101 AWLILWRDNT DNFAILLMLLA AFTVFQYAYA VSGDLNLLRA QVHLNMAAVM
151 FVSVRVSVLL GTETLKECR L KDPVFIPNVI YKNIATITLL HAAAEWLWLEA
201 QTAGFTALAV GFILLAKLRE LHHHELLRKH YVRTYLLQL FAAAGYLWGT
251 AAKLQNLPAF APLHLITLGG MTGVMVMVWL TAGLWHSGFT KLDYFKLCRI
301 AVSILFASAV SRAVLMNVNP IFFITVPEIL TAAVFMLYLL TEVPIFRANA
351 FTDDEP*

```

ORF130ng-1 and ORF130-1 show 92.4% identity in 357 aa overlap:

```

35 orf130-1.pep MRPFFVGAVALAILGALVFFINPGAIVLHRQIFLEMLPAAYGGFLTAALLDWTGFSGNL
orf130ng-1 MRPFFVGAVALAILGALVFFINPGAII LHRQIFLEMLPAAYGGFTTALLDRTGFSGNL
40 orf130-1.pep KPVATLMAALLAASAILPFSPTASFFVAAYVLVLLFCARLWLDRNTDNFALLMLLA
orf130ng-1 KPAATLMAVLLLVAAVLLPFLPQLAAFFVAAYVLVLLFCARLWLDRNTDNFALLMLLA
45 orf130-1.pep AFTVFQYAYAVSGDLNLLRAQVHLNMAAVMFVSVRVSI LLGAELKECR LKDPVFIPNVI
orf130ng-1 AFTVFQYAYAVSGDLNLLRAQVHLNMAAVMFVSVRVSVLLGTEFLKECR LKDPVFIPNVI
50 orf130-1.pep YKNIATITLLHAAAEWLWLEAQTAGFTALAVGFILLAKLRELHHHELLRKH YVRTYLLQ
orf130ng-1 YKNIATIT-LLHAAAEWLWLEAQTAGFTALAVGFILLAKLRELHHHELLRKH YVRTYLLQ
55 orf130-1.pep LFPAAGYLTGAALKQNLPAAPLHLITLGGMTGVMVMVWLTAGLWHSGFTKLDYFKLCR
orf130ng-1 LFPAAGYLTGAALKQNLPAAPLHLITLGGMTGVMVMVWLTAGLWHSGFTKLDYFKLCR
orf130-1.pep IAVPILFAAAVSR AFLMNVNPIFFITVPEILTA AVFVLYLFTFPIFRANAFTDDPEX
orf130ng-1 IAVSILFASAVSRAVLMNVNPIFFITVPEILTA AVFVLYLFTFPIFRANAFTDDPEX

```

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

5
101 COGGCTGGTG TAAGCCGAGA AAACCTGCCG CCATCGATTT TTGGGATATT
151 GCGGCGGAGA GTCCCTCGTC TTTAGAGGAC TACGAGATAC CGCTTTCAGA
201 CGGCATTCGT TCCTCAGGG CAAACGATA TGAATCCGCA CAACATCTT
251 ACTTTTACAG GAANATAGGG AAGTTTGAAG CCTCGGGTT GGATTGGGT
301 ACGCGTGACG GCAAACCTTT GATTGAGACG TTCAACACAG AAGTTTGTGA
351 TTGTTTGAAG AAGCAGGGGT TCGGGCGCAA CGGTCTGTCC GAGCGCGTCC
401 GATGTAA

This encodes a protein having amino acid sequence <SEQ ID 860>:

10
1 MEIRAIKYTA MAALLAFTVA GCRLAGWYEC SSLSGWCKPR KPAIDFWDI
51 GGESPPSLED YEIPLSDGNSR SVRANEYESA QQSYFYRKIG KFEACGLDWR
101 TRDGKPLIET FKQEGFDCLK KQGLRRNGLS ERVRW*

ORF131a and ORF131-1 show 97.0% identity in 135 aa overlap:

15
20
25
orf131a.pep MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLSGWCKPRKPAIDFWDIGGESPPSLED
orf131-1 MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLSGWCKPRKPAIDFWDIGGESPPSLED
orf131a.pep YEIPLSDGNSRVRANEYESAQQSYFYRKIGKFEACGLDWRTRDGKPLIETFKQEGFDCLK
orf131-1 YEIPLSDGNSRVRANEYESAQQSYFYRKIGKFEACGLDWRTRDGKPLIETFKQEGFDCLK
orf131a.pep KQGLRRNGLSERVRWX
orf131-1 KQGLRRNGLSERVRWX

25 Homology with a predicted ORF from *N.gonorrhoeae*

ORF131 shows 89.3% identity over 121 aa overlap with a predicted ORF (ORF131ng) from *N.gonorrhoeae*:

30
35
orf131.pep MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLSGWCKPRKPAIDFWDIGGESPPSLED 60
orf131ng MEIRIVIKYTATAALFAFTVAGCRLAGWYECSSLSGWCKPRKPAIDFWDIGGESPPSLED 60
orf131.pep YEIPLSDGNSRVRANEYESAQQSYFYRKIGKFEACGLDWRTRDGKPLIETFKQGGFDCL 120
orf131ng YEIPLSDGNSRVRANEYESAQQSYFYRKIGKFEACGLDWRTRDGKPLIVERFKQEGFDCL 120
orf131.pep K 121
orf131ng KQGLRRNGLSERVRW 134

40 A complete length ORF131ng nucleotide sequence <SEQ ID 861> was predicted to encode a protein having amino acid sequence <SEQ ID 862>:

1 MEIRIVIKYTA TAALFAFTVA GCRLAGWYEC LSLSGWCKPR KPAIDFWDI
51 GGESPPSLED YEIPLSDGNSR SVRANEYESA QQSYFYRKIG KFEACGLDWR
101 TRDGKPLVER FKQEGFDCL KQGLRRNGLS ERVRW*

Further work revealed the following gonococcal DNA sequence <SEQ ID 863>:

45
50
1 ATGGAATTC GGGTAATAAA ATATACGGCA ACGGCTGCGT TGTTCGATT
51 TACGGTTGCA GGCTGCGGCG TGGCGGGGTG GTATGAGTGT TCGTCCTTGT
101 COGGCTGGTG TAAGCCGAGA AAACCTGCCG CCATCGATTT TTGGGATATT
151 GCGGCGGAGA GtcgcgtGTC TTTAGAGGAC TACGAGATAC CGCTTTCAGA
201 CGGCAATCGT TCCGTCAGGG CAAACGATA TGAATCCGCG CAAAAATCGT
251 ACTTTTATAG GAAAAATAGG AAGTTTGAAG CCTCGGGTT GGATTGGCGT
301 ACGCGTGACG GCAAACCTTT GTTTGAGAGG TTCAACACAG AAGTTTTCGA
351 CTGTTTGGAA AAGCAGGGGT TCGGGCGCAA CGGCCTGTCC GAGCGCGTCC
401 GATGATA

This corresponds to the amino acid sequence <SEQ ID 864; ORF131ng-1>:

```

1 MEIRVIKYTA TAALFAFTVA GCRLAGWYEC SSLSGWCKPR KPAADFDWI
51 GGESPLSLED YEIPLSDGNR SVRANEYESA QKSYFYRKIG KFEACGLDWR
101 TRDGKPLVER FKQEGFDCLE KQLLRNGLS ERVRW*

```

ORF131ng-1 and ORF131-1 show 92.6% identity in 135 aa overlap:

```

5 orf131ng-1.pep MEIRVIKYTATAALFAFTVAGCRLAGWYECSSLSGWCKPRKPAADFDWDIGGESPLSLED
  |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
orf131-1 MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLTWGCKPRKPAADFDWDIGGESPPSLGD

orf131ng-1.pep YEIPLSDGNRSVRANEYESAQKSYFYRKIGKFEACGLDWRTRDGKPLVERFKQEGFDCLE
  |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
10 orf131-1 YEIPLSDGNRSVRANEYESAQKSYFYRKIGKFEACGLDWRTRDGKPLIETFKQGGFDCLE

orf131ng-1.pep KQGLRRNGLSERVWXX
  |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
15 orf131-1 KQGLRRNGLSERVWXX

```

Based on the presence of a predicted prokaryotic membrane lipoprotein lipid attachment site, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 102

20 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 865>

```

1 ATGAACACACA TCCATATTAT CGGTATCGGC GGCACGTTTA TGGCGGGGCT
51 TGCOCGCATT GCCAAGAAG CGGGGTTTGA AGTCAGCGGT TGCGACGCGA
101 AGATGTATCC GCCGATGAGC ACCCAGCTCG AAGCCTTGGG TATAGACGTG
25 151 TATGAAGGCT TCGATCGCGC TCGATTGGAC GAATTTAAG CCGACGTTTA
201 CGTTATCGGC AATGTCGCCA AGCGCGGGAT GGATGTGGTT GAAGCGATT
251 TGAACCTCGG CCGTGCCTAT ATTTCGGCC CGCAATGGCT GTCCGAAAAC
301 GTGCTGCACC ATCATTGGGT ACTCGGTGTG GCGGGGACgC ACGGCAAAAC
351 GACCACCGCC TCCATGCTCG CATGGGTCTT GGAATATgCC GGCCCTCGCGC
401 CGGGCTTCCT TATTGGCGGC GTACC.GGAA AATtTCGGCG TTTCGCGCCG
30 451 CCGTCGCGAA ACGCGCGGCC AAGACCCGAA CAGCCAATCG CCGTTTTTGG
501 TCATCGAAGC CGACGAATAC GACACCGCCT TTTTCGACAA ACGTTCTTAA
551 TtCGTGcATT ACCGTCCGCG TACCGCCGTG TTGAACATC TCGAATTCGA
601 CCACGCGCAC ATCTTTGGCG ACTTGGGGCG GATACAGACC CAGTTCCACT
651 ACTCGTCGCG TACCGTSCCG TCTGAAGCTT TAATCGTCTG CACCGGACCG
35 701 CAGCAAGCGC TCGAAGATAC TTTGACAAA GCCTGCTGGA CGCCGTTGGA
751 AAAATTTCGGC ACGGAACACG GCTGSCA..

```

This corresponds to the amino acid sequence <SEQ ID 866; ORF132>:

```

1 MKHIHIIGIG GTFMGLAAI AKEAGFEVSG CDKMYPPMS TQLEALGIDV
51 YEGFDAQLD EFKADYVIG NVAKRGMDVV EAINLGLPY ISGPQWLSN
40 101 VLHHHVVVLG AGTHGKTTTA SMLAWVLEYA GLAPGFLIG VXGKFRFRFP
151 PAANAAPRPE QPIAVERHRS RRIHRHLFRQ TFXIRALPSA YRRVEQSGIR
201 PRRHLCRLRG TDDEVLPRA YRAVXRLNRL QRTAAKPARY EQRLLDAGG
251 KIRHGTRLA..

```

Further work revealed the complete nucleotide sequence <SEQ ID 867>:

```

45 1 ATGAACACACA TCCATATTAT CGGTATCGGC GGCACGTTTA TGGCGGGGCT
51 TGCOCGCATT GCCAAGAAG CGGGGTTTGA AGTCAGCGGT TGCGACGCGA
101 AGATGTATCC GCCGATGAGC ACCCAGCTCG AAGCCTTGGG TATAGACGTG
50 151 TATGAAGGCT TCGATCGCGC TCGATTGGAC GAATTTAAG CCGACGTTTA
201 CGTTATCGGC AATGTCGCCA AGCGCGGGAT GGATGTGGTT GAAGCGATT
251 TGAACCTCGG CCGTGCCTAT ATTTCGGCC CGCAATGGCT GTCCGAAAAC
301 GTGCTGCACC ATCATTGGGT ACTCGGTGTG GCGGGGACgC ACGGCAAAAC
351 GACCACCGCC TCCATGCTCG CATGGGTCTT GGAATATGCC GGCCCTCGCGC
401 CGGGCTTCCT TATTGGCGGC GTACC GGAAA ATTTCGGCGT TTCCGCGCGC
50 451 CTGCGCGAAA CCGCGCGCCA AGACCCGAAC AGCCAATCGC CGTTTTTCGT
55 501 CATCGAAGCC GACGAATACG ACACCGCCTT TTTTCGACAA CGTTCTTAAT

```

5 551 TCGTGCATTA CCGTCCGCGT ACGCGCGTGT TGAACAATCT GGAATTCGAC
 601 CACGCCGACA TCTTTGCCGA CTGGGCGCG ATACAGACCC AGTTCDACTA
 651 CTTGGTGGCT ACGGTGGCT CTGAAGGCTT AATCGCTTGC AACGGACGGC
 701 AGCAAAAGCT GCAGATACT TTGGACAAG GCTGCTGAC CCGGTTGGA
 751 AAATTCGGCA CGGAACACGG CTGGCAGGCC GCGCAAGCCA ATGCCAGGAA
 801 CTGGTTGACG GTGTTGCTCG ACGCAAAAC CCGCGGACGC GTCAATTCGG
 851 ATTTGATGGG CAGGCACAAC CGCATGAACG CGTCCGCGT CATTTGCGGC
 901 GCGCGTCATG TCGGTTGTCG TATTCAGACC GCGTCCGAAG CCTTGGGCGC
 951 GTTTAAAACG GTCAACCGCC GGATGGAAT CAAGGCAAG CCAACCGGCA
 10 1001 TCACCGTTTA CGACGACTTC GCCCACCACC CGACCGGCAT CGAAACCAAG
 1051 ATTCAGGTTT TCGCGCAACG CGTGGGCGGC CGCGGCTCC TCGCGTCTCT
 1101 CGAACCGGCT TCCACACGTA TGAAGCTGG CACGATGAAG TCCGCCCTCG
 1151 CTGTAAGCCT CAAGAAGGCC GACCAAGTGT TCTGCTACGC CGCGCGCGTG
 1201 GACTGGGACG TCGCCGAAGC CCTCGCGCCT TTGGGCGGCA GGCTGAACGT
 1251 CGGCAAGAC TTGATGCGCT TCGTTGCCGA AATCGTGAAA AACGCGAAG
 1301 TAGGCGACCA TATTTTGGTG ATGAGCAACG GCGGTTTCGG CGGAATACAC
 1351 GGAAAGCTGC TGAAGCTTT GAGATAG

This corresponds to the amino acid sequence <SEQ ID 868; ORF132-1>:

20 1 MKHHIIGIG GTFMGLAAI AKEAGFEVSG CDARKYPPMS TQLEALGIDV
 25 51 YEGFDAQLD EFKADYVIG NVAKRGMDVV EAILNLGLPY ISGPQLSEN
 101 LVHHHWWVLV AGTHGKTTA SMLAWVLEYA GLAPFLIGG VPENRGVSAR
 151 LPQTERPDEN SQSFFVIEA DEYDTAFDPK RSKFVHYRER TAVLNLEFD
 201 HADIFADLGA IQTFHYLVR TVPSEGLIVC NGRQSLQDT LDKGCWTFVE
 251 KFGTEHGWA GEANADGSFD VLLDKTAGR VRWDLGRHN RMNALAVIAA
 301 ARHVGVDIQT ACEALGAFKN VKRMEIKGT ANGITYYDDF AHHPITAET
 351 IQGLRQRVGG ARILAVLEPR SNTMKLGTMK SALTPLSKA DQVFCYAGGV
 401 DWDVAEALAP LGRLNVGKD FDAFVAEIVK NAEVGDHILV MSNGGPGGIH
 451 GKLEALR*

Computer analysis of this amino acid sequence gave the following results:

30 Homology with the hypothetical o457 protein of *E. coli* (accession number U14003)

ORF132 and o457 show 58% aa identity in 140 aa overlap:

35 Orf132: 4 IHHIIGIGGTFMGLAAIAKEAGFEVSGCDARKYPPMSTQLEALGIDVYEGFDAQLDEFK 63
 IHI+GI GTFMGLA +A++ G EV+G DA +YFPMST LE GI++ +G+DA+QL+ +
 o457: 3 IHHILGCGTFMGLAMLARQLGHEVTGSDANVYFPMSTLLEKQGIELIQYDASQLEP-Q 61
 Orf132: 64 ADVVYIGNVAKRGMDVVEAILNLGLPYISGPQLSENVLHHHWWVLGVAGTHGKTTASML 123
 D+ +IGN RG VEA+L +PY+SGPQL+ VL +VL VAGTHGKTTA M
 o457: 62 PDVVIIGNAMTRGNPCVEAVLEKNIPYMSGPQWLHDVFLDRWVLAAGTHGKTTAGMA 121
 40 Orf132: 124 AWLEYAGLAPGFLIGGVXG 143
 W+LE G PGF+IGGV G
 o457: 122 TWILEQCQYKPGFVIGGVPG 141

Homology with a predicted ORF from *N. meningitidis* (strain A)

45 ORF132 shows 74.6% identity over a 189aa overlap with an ORF (ORF132a) from strain A of *N. meningitidis*:

50 orf132.pep MKHHIIGIGGTFMGLAAIAKEAGFEVSGCDARKYPPMSTQLEALGIDVYEGFDAQLD
 orf132a MKHHIIGIGGTFMGLAAIAKEAGFEVSGCDARKYPPMSTQLEALGIDVYEGFDAQLD
 55 orf132.pep EFKADYVIGNVAKRGMDVVEAILNLGLPYISGPQLSENVLHHHWWVLGVAGTHGKTTA
 orf132a EFKADYVIGNVAKRGMDVVEAILNLGLPYISGPQLSENVLHHHWWVLGVAGTHGKTTA
 130 140 150 160

-491-

```

orf132.pep  SMLAWVLEYAGLAPGFLIGGVXGKFR---RFRPPAANAAPRFEQFI-----AVFR
            ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
orf132a     SMLAWVLEYAGLAPGFXIGGVFNFSVSARL-PQTRPQDPNSQSPFFVIEADEYDTAFDD
            130      140      150      160      170

5           170      180      190      200      210      220
orf132.pep  HRSRRIRHRLFRQTFXIRALPSAYRRVQSGIRPRHLRLGRDTPVPLPRAYRAVXRRL
            :||: ||:|
orf132a     KRSKFVHYRPRTAVLNNLEFDHADIFADLGAIQTFHHLVRTVPSSEGLIVCNCRQSLQD
            180      190      200      210      220      230

```

The complete length ORF132a nucleotide sequence <SEQ ID 869> is:

```

1  ATGAACACAC  TCCACATTAT  CGGTATCGGC  GGCACGTTTA  TGGGTGGGAT
5  TGGCGCCATT  GCCAAAGAAG  CAGGCTTTGA  ANTCAAGCGT  TGGCATGGGA
10  AGATGTATCC  GCGCATGAGC  ACCAGCTCG  AAGCCTTGG  CATAGCGCTG
15  TATGAAGCTT  TGACACCGCT  CAGGTTGAG  GAATTTAAG  CGACGCTTTA
20  CGTTATCGGC  AATGTCGCG  AGCGCGGAT  GAGTGTGGT  GAAGCGATT
25  TGAACGTCAC  ATCGCTTAT  ATTTCCGCG  CGCATGGCT  GGCTGAAAC
30  NTGCTGCAC  ATCATTGNN  ACTCGGCTG  GCGNGAAGC  GCGCAAAAC
35  GACACCGCG  TCTATGCTG  CGTGGGCTT  GGAATATGC  GGACTCGAC
40  CGGCTTCNT  TATCGCGCG  GTACCGGAA  ACTTCAGCT  TTCGCGCGC
45  CTGCGCACA  CGCGCGCCA  AGACCGGAA  AGCCAATCG  CGTTTTCTG
50  CATTGAAGCC  GAGCAATAC  ACACCGCGT  TTTGCACAA  GCGTCAAA
55  TCGTGCAAT  CGTCCGCGT  ACCGCGCTG  TGAACAATC  GGAATTCG
60  CACGCGACA  CTTTCGCGA  TTTGGCGCG  ATACGACCC  AGTTCCACA
65  CCTCGTCTG  ACCGTGCGT  CTGAAGCCT  CATCGTCTG  AACGGACGC
70  AGCAAAAGC  CAGAACACT  TTGACAAAG  CTGCTGGAC  GCGGCTGAA
75  AAATTCCGCA  CGGAACACG  CTGCGAGCC  GCGCAAGCA  ATGCGGATG
80  CTGCTTGC  GTGTGCTTG  AOGCGAAAA  AGCGGACAC  GTGCTTTGA
85  GTTTGATGG  CGGACACAC  CGCATGAAG  CGCTCGNGT  CATGCGCGC
90  GCGGCTCAT  CGGAGTNGA  CATTGACG  GCGTGGAG  CTTTGGAC
95  GTTTAAAGC  GTCGAACGC  GCATGGAAT  CAAAGCAG  GCAACGTA
100  TCCCGTTT  CACAGCTTC  CGCCGCTAT  CGACCGCTT  CGAACACG
105  ATTCAAGTT  TCGCGCAGG  GCTCGCGGC  GCGCGATCC  TCGCGCTCT
110  CGAACGCGT  TCCATACGA  TGAAGCTGG  TACGATGAA  GCGCGCTCG
115  CGCAAGCCT  CAAAGAAGC  GACCAAGTG  TCTGNTACG  CGGCGCGCG
120  GACTGGGAC  TTGCGGAAG  CCTCGCGCT  TTGGCGGCA  GGCTGCACT
125  CGGCAAGAC  TTGATGCTT  TCGTTCGCA  AATCGTGAA  AACGCGAAG
130  CAGGCGACA  TATTTTGGT  ATGAGCAAG  GCGGTTTCG  CGGAATAC
135  ACCAAACTG  TGGACGCTT  GAGATAG

```

This encodes a protein having amino acid sequence <SEQ ID 870>:

```

1  MKHIHIIGIG  GTFMGGIAAI  AKEAGFEKSG  CDKMYPPMS  TQLEALGIV
5  YEGFDTAQLD  EFKADVYVIG  NVAKRGMDV  EAILNRGLPY  ISGPQWLAE
10  XLHHHWXWLG  AXTHGKTTTA  SMLAWVLEYA  GLAPGXFIGG  VPENFVSAR
15  LPQTFRQDPN  SQSPFFVIEA  DEYDTAFDD  RSKFVHYRFR  TAVLNNLEFD
20  HADIFADLGA  IQTFHHLVLR  TVPSEGLIVC  NGRQSLQDT  LDRGCTVFE
25  KFTGTEHGWA  GERADGSGFD  VLLDGKKAGH  VWMLVGGHIN  RMLALVTA
30  ARHAGVDTCT  ACEALSTVTS  VRRMEIKGT  ANGITYDDF  AHHPTAETET
35  TQGLRGVYGG  ARTLAVLEFR  SNTMKLGTMK  AALPASLKEA  DOVFXAYAGA
40  DWDVAEALAP  LGRRLHVSKD  FDFVFAEIVK  NAEAGDHILV  MSNGGFGGIH
45  TKLLDALR*

```

ORF132a and ORF132-1 show 93.9% identity in 458 aa overlap:

```

orf132a.pep  MKHIHIIGIGGTFMGGIAAIAKEAGFEKSGCDKMYPPMSTQLEALGIVGVEGFDTAQLD
            ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
orf132-1     MKHIHIIGIGGTFMGGIAAIAKEAGFEKSGCDKMYPPMSTQLEALGIDVYEGFDAAQLD
            55

orf132a.pep  EFKADVYVIGNVAKRGMDVVEAILNRGLPYISGPQWLAEAXLHHHWXWLGVAETHGKTTTA
            ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
orf132-1     EFKADVYVIGNVAKRGMDVVEAILNLGLPYISGPQWLSENVLHHHWXWLVAGTHGKTTTA
            60

orf132a.pep  SMLAWVLEYAGLAPGFXIGGVFNFSVSARLPQTRPQDPNSQSPFFVIEADEYDTAFDDK
            ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
orf132-1     SMLAWVLEYAGLAPGFLIGGVFNFSVSARLPQTRPQDPNSQSPFFVIEADEYDTAFDDK

```

	orf132a.pep	RSKFVHYRPRTAVLNNLEFDHADIFADLGAIQTQFHLVVRTVSEGLIVCNGRQQSILQDT
	orf132-1	RSKFVHYRPRTAVLNNLEFDHADIFADLGAIQTQFHYLVVRTVSEGLIVCNGRQQSILQDT
5	orf132a.pep	LDKGCWTPVEKFGTEHGWQAGEANADGSFDVLDDGKKAGHVANSIMGGHNRMNALAVIAA
	orf132-1	LDKGCWTPVEKFGTEHGWQAGEANADGSFDVLDDGKTAGRVKDWLMGRHNRMNALAVIAA
10	orf132a.pep	ARHAGVDIQTACEALSTFKNVKRRMEIKGTANGITVYDDFAHHPPTAIETTIGQLRQRVGG
	orf132-1	ARHVGVDIQTACEALGAFKNVKKRMEIKGTANGITVYDDFAHHPPTAIETTIGQLRQRVGG
15	orf132a.pep	ARILAVLEPRSNMKGKLTMAKALPASLKEADQVFXVYAGGADWDVAEALAPLGGRLHVNGKD
	orf132-1	ARILAVLEPRSNMKGKLTMAKALPVSLKEADQVFCYAGGVDWDVAEALAPLGGRLHVNGKD
20	orf132a.pep	FDAFVAEIVKNAEAGDHILVMSNGGFGGIHTKLLDALRX
	orf132-1	FDAFVAEIVKNAEAGDHILVMSNGGFGGIHGLLEALRX

Homology with a predicted ORF from *N.gonorrhoeae*

ORF132 shows 89.6% identity over 259 aa overlap with a predicted ORF (ORF132ng) from *N. gonorrhoeae*:

25	orf132.pep	MKHIHIIGIGGTFMGGLAAIAKEAGFEVSGCDAMKYPMPSTQLEALGIDVVEYEGFDAALQD	60
	orf132ng	MKHIHIIGIGGTFMGGLAAIAKEAGFEVSGCDAMKYPMPSTQLEALGIDVVEYEGFDAALQLE	60
30	orf132.pep	EFKADVYVIGNVAKRGMDDVEAILNLGLPYISGQWLSENVLHHHWLVGAGTHGKTTTA	120
	orf132ng	EFQADIYVIGNVARRGMDDVEAILNRLGLPYISGQWLSENVLHHHWLVGAGTHGKTTTA	120
35	orf132.pep	SMLAWVLEYAGLAPGLIGGVXGKFRFRFPFAANAAPRPEQPIAVFHRSRIRHRLFRQ	180
	orf132ng	SMLAWVLEYAGLAPGLIGGVPKFRFRFPPTANAASRPEQPIAVFHRSRIRHRLFRQ	180
40	orf132.pep	TFXIRALPSAYRRVEQSGIRPRHLRLGRDTPFVLPFRAYRAVXRLNRLQRTAAKPARV	240
	orf132ng	TLQIRALPSAYRRVEQSGIRPRHLRLGRDTPFVPPRAHRTIRPRHLRLQRTAAKPARV	240
45	orf132.pep	FGQRLLDAGGKIRHRTLRAD	259
	orf132ng	FGQRLLDAGGKIRHRTLRADW	261

An ORF132ng nucleotide sequence <SEQ ID 871> was predicted to encode a protein having amino acid sequence <SEQ ID 872>:

45	1	MKHIHIIGIG	GTFMGGLAAI	AKEAGFKVSG	CDAMKYPFMS	TQLEALGIGV
	51	HEGFDAALQLE	EFQADIYVIG	NVARRGMDDV	EAILNRGLPY	ISGFWLAEIN
	101	VLHHHWLVG	AGTHGKTTTA	SMLAWVLEYA	GLAPGLIGG	VPGKFRFRFP
	151	PTANAASRPE	QQIAVFRHRS	RRIRHRLFRQ	TLQIRALSPA	YRRVEQSGIR
50	201	PRRHLRLRGR	DTDVPVPPRA	HRTIRPRHRL	QRTAAKPARV	FGQRLLDAGG
	251	KIRHRTLRAD	W*			

Further work revealed the following gonococcal DNA sequence <SEQ ID 873>:

55	1	ATGAACACA	TCCACATTAT	CGGTATCGGC	GGCACGTTTA	TGGGCGGGAT
	51	TGCGCCGATT	GCACAAGAGC	AGCGGTTTCA	AGTCAGCGGT	TGCGACGCGAT
	101	AGATGTATCC	GCCGATGAGC	ACCACGCTCG	AGCGCTTGGG	CATAGCGCTA
	151	CACGAAGGCT	TCGATCGCGC	CGAGTTGGAA	GAATTTCAAG	CCGATATTTA
	201	CGTCATCGCG	AATGTCGCCA	GGCGCGGGAT	GGATGTGGTC	GAGGCGATT
	251	TGAACCGCTG	GCTGCCTTAT	ATTTCCGGCC	CGCAATGGCT	GGCTGAAGAA
	301	GTGCTgcacc	atcatTGGgt	ACTCGGCGTG	GcagggaCGC	ACGGcaaaAc
	351	gaccaCgcGc	tCCATGCTCG	CCTGGGTCTT	GGAATATGCG	GGACTCGCGC
60	401	CGGGCTTCTT	CATCGGCGGT	gtaccggaAA	ATTTCCGGCT	TCCGCGCGCT

5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
95
100
105
110
115
120
125
130
135

CTACCGCAAA CGCGCGCTCA AGACCCGAAC AGCAAATGCG CGTTTTTGGT
CATCGAAGCC GACGAATACG ACACCCGCTT TTTGACAAA CGCTCCGAAT
TCGTGCATTA TCGCCGCGCT TGAACAATCT GGAATTCGAC
CACGCGGACA TCTTCCGCGA CTTGGGCGCG ATACAGACCC AGTTCCACCA
CCTCGTGCGC ACGTACCAT CCGAAGGCCT CATGCTGPG AACGGACAGC
AGCAAAGCCT GCAAGTACT TTGGACAAG GCTGCTGGAC CCGGCTGAA
AAATTCGGCA CGGACACGCG CTGGCAGATT GGTGAAGTCA ATGCCAGCGG
CTCGTTTCGAC GTATTGCTTG ACGGCAAAA AGCCGGACAC GTGCGATGGG
ATTTGATGGG CGGACACAAC GCGATGAACG CGCTGCGCGT CATGCTGCC
GCACGCCATC CCGGAGTCGA TGTTCAGACG GCCTGCGAAG CCTTGGGTGC
GTTTAAAAAC GTCAACGCC GCATGGAAT CAAAGAGCG CCAACCGGCA
TCACCGTTTA CGACGATTC GCCCACCACC CGACCGCCAT CGAACAACAG
ATTCAGGTT TCGGCCAACG TGTGCGCGCG CGCGGCATCC TCGCGCTCT
CGAGCGCGCT TCCAACCA TGAACCTCG CACGATGAAG TCCGCCCTCG
CCGCAAGCCT CAAAGAAGCC GACCAAGTGT TCTGCTACGC CGCGGCGCG
GACTGGGACG TTGCGGAAGC CCGCGCGCT TTTGGCTGCA GGCTGGCGGT
CGGTAAGAT TTCGATACCT TCGTGGCGCA AATGTGAAA AACGCCCGAA
CGCGGACCA TATTTTGGT ATGACACAG CGGGTTCGG CGGAATACAC
ACCAAATGCG TGGACGCTT GAGATAG

This corresponds to the amino acid sequence <SEQ ID 874; ORF132ng-1>:

1 MKHIHIGIG GTFMGGIAI AKEAGFKVSG CDAMYPFMS TQLEALGIVG
51 HEGFDAAQLE EFQADIYVIG NVARRGMDV FAINRGLPY ISGPQWLAEN
101 VLHHHWLVG AGTHGKTTTA SMLAWVLEYA GLAPGLIG VPENFGV SAR
151 LPQTPRQDPN SKSPFFVIEA DEYDTAFDDK RSKFVHYRFR TAVNNLEFD
201 HADIFADLGA IQTFHHLVR TVPSEGLIVC NGQQSLQDT LDGKCWTVE
251 KFGTGHGWQI GEVNADGSFD VLLDGKAGH VAWDLMGHNN RMNALAVIAA
301 ARHAGVDVQT ACEALGAFKN VKRMEIKGT ANGITYDDF AHHPATAIET
351 IQGLRQRVGG ARILAVLEPR SNTMKLGTMK SALPASLKEA DQVFCYAGGA
401 DWDVAEALAP LGCRILRVGKD FDTFVAEIVK NARTGDHILV MNGGFGGHI
451 TKLLDALR*

ORF132ng-1 and ORF132-1 show 93.2% identity in 458 aa overlap:

orf132ng-1.pep MKHIHIGIGGT FTMGGIAIAKEAGFKVSGCDAMYPFMS TQLEALGIVGHEGFDAAQLE
orf132-1 MKHIHIGIGGT FTMGGIAIAKEAGFEVSGCDAMYPFMS TQLEALGIDVYEGFDAAQLD
orf132ng-1.pep EFQADIYVIGNVARRGMDVVEAILNRLGLPYISGPQWLAENVLHHHWLVGAGTHGKTTTA
orf132-1 EFKADVYVIGNVAKRGMDVVEAILNRLGLPYISGPQWLSENVLHHHWLVGAGTHGKTTTA
orf132ng-1.pep SMLAWVLEYAGLAPGLIGGVPENFVGSARLPQT PRQDPNPSKSPFFVIEADEYDTAFDDK
orf132-1 SMLAWVLEYAGLAPGLIGGVPENFVGSARLPQT PRQDPNPSKSPFFVIEADEYDTAFDDK
orf132ng-1.pep RSKFVHYRPR TAVNNLEFDHADIFADLGA IQTFHHLV R TVPSEGLIVCNGQQSLQDT
orf132-1 RSKFVHYRPR TAVNNLEFDHADIFADLGA IQTFHHLV R TVPSEGLIVCNGQQSLQDT
orf132ng-1.pep LDGKCWTVEKFGTGHGWQIGEVNADGSFDVLLDGKAGHVAWDLMGHNNRMNALAVIAA
orf132-1 LDGKCWTVEKFGTGHGWQIGEVNADGSFDVLLDGKAGHVAWDLMGHNNRMNALAVIAA
orf132ng-1.pep ARHAGVDVQTACEALGAFKNVKRMEIKGTANGITYDDFAHHPATAIETT IQGLRQRVGG
orf132-1 ARHAGVDVQTACEALGAFKNVKRMEIKGTANGITYDDFAHHPATAIETT IQGLRQRVGG
orf132ng-1.pep ARILAVLEPR SNTMKLGTMKSALPASLKEADQVFCYAGGADWDVAEALAPLGCRILRVGKD
orf132-1 ARILAVLEPR SNTMKLGTMKSALPASLKEADQVFCYAGGADWDVAEALAPLGCRILRVGKD
orf132ng-1.pep FDTFVAEIVKNARTGDHILVMNSGGFGGHIHKKLLDALR
orf132-1 FDAFVAEIVKNAEVDHILVMNSGGFGGHIHKKLLDALR

In addition, ORF132ng-1 is homologous to a hypothetical *E. coli* protein:

pir||S56459 hypothetical protein o457 - Escherichia coli >gi|537075 (U14003)
 ORF_o457 [Escherichia coli] >gi|1790680 (AE000494) hypothetical 48.5 kD protein
 in fbp-pmba intergenic region [Escherichia coli] Length = 457

Score = 474 bits (1207) Expect = e-133

Identities = 249/439 (56%), Positives = 294/439 (66%), Gaps = 13/439 (2%)

Query: 22 KEAGFKVSGCDKMYPPMSTOLEALGIGVHEGFDAALQEEFQADIYVIGNVARRGMQDVE 81
 ++ G +V+G DA +YPPMST LE GI + +G+DA+OLE Q D +IGN RG VE
 Sbjct: 21 RQLGHEVTGSDANVYPPMSTLLEKQGLIQQGYDAQLEP-QPDLVIIGNMTRGNPCVE 79

Query: 82 AILNRGLPYISGPQWLAEVLHHHVVVLGAGTHGKTTTASMLAWVLEYAGLPGFLIGGV 141
 A+L + +PY+SGPQWL + VL VVL VAGTHGKTTTA M W+LE G PGF+IGGV
 Sbjct: 80 AVLEKNIPYMSGPQWLHDFVLRDRWVLA VAGTHGKTTTAGMATWILEQCGYKPGFVIGGV 139

Query: 142 PENFVGVSARLEQTFRQDPNSKSPFFVIEADEYDTAFFDKRSKFVHYRPRTAVLNNLEFDH 201
 P NF VSA L +S FFVIEADEYD AFFDKRSKFVHY PRT +LNNLEFDH
 Sbjct: 140 FGNFVSAHL-----GESDFFVIEADEYDCAFFDKRSKFVHYCPRTLILNNLEFDH 190

Query: 202 ADIFADLGAIQOTFHHLVRTVPSEGLIVNCQQQSQDTLDKGCWTPVEKPGTGHGQIG 261
 ADIF DL AIQ QFHHLV RP +G I+ +L+ T+ GCW+ E G WQ
 Sbjct: 191 ADIFDDLKAIQKQFHHLVRIVPGQGRIIWFPENDINLKQTMAMGCWSEQLVGEQGHGQAK 250

Query: 262 EVNADGS-FDVLDDGKAGHVAVMDLGGHNRNMALAVIAARHAGVDVQTACEALGAFKN 320
 ++ D S ++VLLDG+K G V W L+G HN N L IAAARH GV A ALG+F N
 Sbjct: 251 KLTTDASEWEVLLDGEKVGEVWKS LVGEHNHMLMAIAARHVGVAPADAANALGSFIN 310

Query: 321 VKRMEIKGTANGITVYDDFAHHPHTAETTTIQLRQRVGG-ARILAVLEPRSNMTKLG 379
 +RR+E++G ANG+TVYDDFAHHPHTAI T+ LR +VGG ARI+AVLEPRSNMTK+G
 Sbjct: 311 ARRLLELRGEANGVTYDDFAHHPHTAILATLAALRGKVGGTARI IAVLEPRSNMTKMGIC 370

Query: 380 KSALPASLKEADQVF-CYAGGADMDVAEALAPLGRCLRVGKDFDTFVAIEVKNARTGDHI 438
 K L SL AD+VF D DT +VK A+ GDHI
 Sbjct: 371 KDDLAPSLGRADEVFLQPAHIFWQVAEAEACVQPAHWSGDVTIADMVVKTAQPGDHI 430

Query: 439 LVMSNGGFGGINTKLLDAL 457

LVMSNGGFGGIN KLLD L

Sbjct: 431 LVMSNGGFGGINTKLLDGL 449

Based on this analysis, it was predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF132-1 (26.4kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 20A shows the results of affinity purification of the His-fusion protein, and Figure 20B shows the results of expression of the GST-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for FACS analysis (Figure 20C) and ELISA (positive result). These experiments confirm that ORF132 is a surface-exposed protein, and that it is a useful immunogen.

Example 103

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 875>

1 ..CCGGGCTATT ACGGCTCGGA TGACGAATT AAGCGGCAT TCGGAGAAAA
 51 CTCGCGCGACA TmCAAGAAAC ATTGCAACCG GAGCTGCGGG ATTATGTAAC
 101 CCGTATGGA AARATACCG AAAAGCGCG CCAACAACCA TCGGTTCAGC
 151 ATTATGTCGG ACTTTCGCGA TTATTTCTAG CCGTTGCGCA GCTATTTCGG
 201 CACACACCGT ATGCCCAACA TCCAGAAGT GTATTTTTC CAAATCGCG
 251 ACTCCGCGCT TCACACCGCC TTAACAACAG GCGCGCAAA CACTTGCAG
 301 TTTGGCTTC ATACATATA AAAAGATTG TTAACAACAG ATGATACATT

351 AGGATTAAAA CTGGTCGGCT ACCGCAGCGG CATCGACAAC TACATCCACA
 401 ACCTTTACCG GAAATGGTGG GATTTTGAAG GGGATATATC GAGCTGGTGC
 451 AGCGATCGCG GCGTGGCTTA CACCATCCAA CATCGCTCAA TCAWAGTAA
 501 AGTGCATCAA nnnnnnnnnn nnnnnnnnnn nnnnTACGAT TATGGGCGTT
 5 551 TTTTCCACCA CCTTCTTCTAC GCCTATCAAA AAGACACGA ACCGACAAC
 601 TTACGCGATG CGAGCGGAATC GCCCAACAAT CGCTCCAAAG AAGACCAACT
 651 CAAACAAGGT TATGGGTGGA CGAGGGTTTC GCCTCGCGC CGAGATTTACG
 701 GACGTTTGGG AGTCGGTACG CGCTGGTTGG GCAACAAACT GACTTTGGGC
 10 751 GGGCGGATGC GCTATTTTCG CAAGAGCATC CGCGGACGG CTGAAGRACG
 801 CTATATCGAC GGCACCAACG GGGGAAATAC CAGCAATTTT CGGCAACTGG
 851 CGAAGCGTTC CATCAACAAA ACCGAAACTC TTGCCCGCCA GCCTTTGATT
 901 TTWGATTTTt ACGCCGCTTA CGAGCGCGAAG AAAAACCTTA TTTTCCGCGC
 951 CGAAGTCAAA AATCTGTTTC ACAGCGCTTA TATCGATCCG CTCGATGGG
 1001 GCAATGATGC GGCAC. GAG CGTTATTACA GCTCGTTTCA CCGAAGACG
 15 1051 AAGGACrRAG ACGTAACGTG TAATGCTGAT AAAACGTTGT GCAACGGCAA
 1101 ATACGGCGGC ACAAGCAAAA GGGTATTGAC CAATTTTGCA CGCGGACGCA
 1151 CCTTTTgAT GACGATGAGC TACAAGTTTT AA

This corresponds to the amino acid sequence <SEQ ID 876; ORF133>:

20 1 . . . PGYYGSDDEF KRAFGENSPT KXKHCNRSCG IYEPVLKKYG KKRANNHSVS
 51 ISAPGDYFM QGTDSRSRTF QIGDSGVHTA LKPERANRVA
 101 FGPYTKKGL LKQDDYGLK LVGRSRIDN YHNVYKRW DLNGLPSRV
 151 SSTGLAYTIQ HRXEXDKVHQ XXXXXXXXYD YGRFPTNLSY AYXKSTOPTN
 201 FSDASESPNN ASKEDQLKQG YGLSRVSLP RDYGRLEVGT RWLGNKLTGL
 25 251 GAMRYFGKSI RATAERYID GTNGNTSNF RQLGKRSIKQ TETLARPLI
 301 XDFNAAYEPK KNLFRAEVK NLFDRRYID LDAGNDAAXE RYSSFDPK
 351 KDXDVTGNAD KTLNCKGYKG TSKSVLNTFA RGRTHLMTMS YK*

Further work revealed the further partial DNA sequence <SEQ ID 877>:

30 1 GAGGCGCAGA TACAGGTTTT GGAAGATGTG CACGTCAAGG CGAAGCGCGT
 51 ACCGAAAGAC AAAAAGATGT TTACCGATGC GCGTGCCTGA TCGACCGCTG
 101 AGGATATATT CAAATCCAGC GAAACCTCGT ACAACATCGT ACGCAGCATC
 151 CCGGTCGCGT TTACACAGCA AGATAAAAGC TCGGGCATTG TGCTTTGAA
 201 TATTCGCGGC GACAGCGGGT TCGGCGGGT CAATACGATG GTGAGCGGCA
 35 251 TCACGCGAGC CTTTTATTTC ACTTCTACCG ATGCGGGCAG GCGAGCGGCT
 301 TCATCTCAAT TCGGTGCATC TGTGACAGC AATTTTATTG CCGGACTTGA
 351 TGTCTGAAA GGCAGCTTCA CGGCTCGGC AGGCATCAAC AGCCTTGGCG
 401 GTTCGCGGAA TCTCGGACT TTAAGCTGG ATGACGCTG TCGAGCAAT
 45 451 AATACTACG GCCTGCTGCT AAAAGTCTG ACCGCAACCA ATTCAACCAA
 501 AGGTAAATGCG ATGCGCGCGA TAGGTGCGCG CAATTTGGCTG GAAAGCGAAG
 551 CATCTGTCGG TGTGCTTTAC GGGCAGACGA GCGCGAGCGT GGCACAAAT
 601 TACCGCGTGG GCGGCGCGGG GCAGCAGCAT GGAATTTTTC GCGCGGAATA
 701 TTTGGAACGG CGCAAGCAGC GATATTTTGT ACAAGAGGGT GCTTTGAAAT
 751 TCAATTCGGA CAGCGGAAAA TGGGAGCGGG ATTTACAAAG GCAACAGCTG
 801 AARATCAAGC CGTATAAAAA TTACAAACAC CRAAGACTAC AaAATACAT
 851 CGRAGAGCAT GACAAAAGCT GGGCGGAAAA CTTg. CaCG CAATACGACA
 45 901 TTACCCCATC CGATCGCTCC AGCCTGAAG AGCAGTCGCG AGGCAATCTG
 901 TTTAAATTTG AATACGAGCG CGTATTCAT AATACAGCG CGCAATTTTC
 951 CGATTTAAAC ACCAAAAATCG CGACGCGCAA AATCATCAAC CGCAATTTATC
 1001 AGTTCAAATTA CGTTTTGCTT TTGAACCCGT ATACCAACCT CCAATGTGAC
 1051 CGAGCOTCAA ATTGCGGCGAG GCAGAAATAT CGGAAGGGT CGAGTTTAC
 1101 AGCTGGGGG ATTTTGAACG CTACACACG GCGAARATC GCGAARATC
 50 1151 TCGACCTCAA CACACCGCC ACCTTCCGCG TGCCCGCGGA AACCAGCTTG
 1201 CAAACCACTT TGGGCTTCAA TTATTTCCAC AACGAATACG GCAAAAGACG
 1251 CTTTCTGAA GAATTTGGGG TGTTTTTCGA CGGTCTGATC CAGGACRACG
 1301 GGGTTTATTC CTATTTGGGG GGGTTTAAAG GCGATAAAGG GCTGCTGCC
 55 1351 CAAAAATCAA CCATTGTCCA ACCGGCGCGC AGCCATAATT TCAACAGCTT
 1401 CTACTTCGAT CGCGCGCTCA AAAAAGACAT TTACCGCTTA AACTACAGCA
 1451 CCATACCGCT CGGCTACCGT TTCGGCGCGC AATATACGGG CTATTACGGC
 1501 TCGGATGAGC AATTTAAGCG GGCATTCGGA GAAACTTCGC CGCATACAAA
 1551 GAAACATTTG AACCGGAGCT CGGGGATTTA TGAACCCGTA TTGAAAAAT
 60 1601 ACGCGAAAAA CGCGGCCAAC AACCATTTCG TCAGCATTAG TGCGCATCTC
 1651 GGGCATTTAT TCATGCGGTT CGCCAGCTAT TCGGCGCACAC ACCGATTTGCC
 1701 CAACATCCAA GAATGTATT TTCCCAAT CCGGGACTCC GGGCTTCACA
 1751 CGCGCTAAA ACGAGCGCGC GCARAACATT GGCATTTGTT CTTCAATTC
 1801 TATAAAAAG GATTTTAAA ACATTTGAT ACATTTGAT TAACTTAACT
 65 1851 CGCTACCGC AGCGCGATCG ACACATACAT CAACACCTGTT TACGGAAAT
 1901 GGTGGGATTT GAACGGGGAT ATTCCGAGCT GGGTCAGCAG CACCGGGCTT

1951	GCCTACACCA	TCCACATCG	CAATTTCAA	GACAAAGTGC	ACAAACACGG
2001	TTTGAGTTG	GAGCTCAAT	ACGATTATGG	GGCTTTTTC	ACCAACCTTT
2051	CTTAGCCCTA	TCAAAAAGC	ACGCAACCGA	CCAACCTCAG	CGATGCGGAC
2101	GAATCGCCCA	ACAATCGCTG	CAAAGAAGAC	CAACTCAAC	AAGGTTATGG
2151	GTTGAGCAGG	GTTTCCGCC	TGCCGCGAGA	TACGGACGT	TTGGAAGTCG
2201	GTACGCGCTG	GTTGGGCAAC	AACTGACTTT	TGGGCGGCGC	GATGCGCTAT
2251	TTCCGCAAGA	GCATCCGCGC	GACGGCTGAA	GAACGCTATA	TCGACGGCAC
2301	CAACGGGGGA	AATACACGCA	ATTTCCGGCA	ACTGGGCAAG	CGTTCATCAT
2351	AACAAACCGA	AACTCTTGCC	CGCCAGCCTT	TGATTTTGA	TTTTTACGCC
2401	GCTTACGAGC	CGAAGAAAA	CCTTATTTTC	CGCGCCGAAG	TCAAAANTCT
2451	GTTTCAGCAGG	CGTTATATCG	ATCCGCTCGA	TCGGGCAAT	GATGCGGCAA
2501	CGCAGCGTTA	TTACAGCTCG	TTGACCCGCA	AAGACAAGGA	CGAAGACGTA
2551	ACGTGTAATG	CTGATAAACC	GTTGTGCAAC	GGCAATACG	CGCGGCAACG
2601	CAAAAGCGTA	TTGACCAATT	TTGACGCGGG	ACGACCTTTT	TTGATGACGA
2651	TGAGCTACAA	GTTTAA			

This corresponds to the amino acid sequence <SEQ ID 878; ORF133-1>:

1	EAQIQVLEDV	HVKAKRVKPD	KKVETDARAV	STRQDIFKSS	ENLDNIVRSI
51	FGAFTQQQKS	SGIVSNIFRG	DSGFGRWNIM	VDCITTFYFS	TSTDAGRAGS
101	SSQFGASVDS	NFIAGLDVVK	GSFSGSAGIN	SLAGSANLR	LGVDVVQVGN
151	NTYGLLLKGL	TGNTSTKNGA	MAAIGARKWL	ESGASVGVLY	GHSRRSVAGN
201	YRVGGGGQHI	GNFGASYLER	RKQRYFVQEG	ALKFNSDSGC	WERDLQROOQ
251	KYKPKYKNYN	QELQKYEIEH	DKSWRENLPX	QYDITPIDPS	SLKQSSAGNL
301	FKLEYDGVFN	KYTAQFRDLN	TKIGSRKIIN	RNYQFNLYGS	LNPVTNLNLT
351	AAVNSGRQKY	PKGSKFTGWL	LKDFETVNN	AKILDNLNNTA	TRFLPRETEL
401	QTLGIFYNFH	NEYGKNRFFE	ELGLFFDGD	QDNGLSYLYG	RFKGDKGLPL
451	QKSTIVQFAG	QYFNTFYFD	AALKKDIYRL	NYSTNTVGYR	EGGEYTYGYG
501	SDDEFKRAFQ	ENSPTYKKHC	NRSCGIEYFV	LKYGKRRAN	NHSVISIADF
551	GDYFMFASY	SRTHRMPIQ	EMYFSQIGDS	GVHTALKPER	ANTWQGFNT
601	YKKGLLKQDD	TLGLKLVGYR	SRIDNYIHNV	YKGVWLNDG	IPSWVSTGL
651	AYTIQHRNFK	DKVHKRGFEL	ELNIDYGRFF	TNLSYAYQKS	TOPTNFSAS
701	ESPNNASKED	QLKQGYGLSR	VSLPRDYGR	LEVGTWRWGN	KLTGGAMRY
751	FGKSIRATAE	ERYIDGTNGG	NISNFRQLGK	RSIKQTETLA	RQPLIFDYFA
801	AYEPKKNLIF	RAEVKMLFDR	RTIDPLDAGN	DAATQRYYS	FDPKDKDEVD
851	TCNADKTLGN	GKYGGTSKSV	LTFNFARGRT	LMTMSYK*	

Computer analysis of this amino acid sequence gave the following results:

Homology with with the probable TonB-dependent receptor HI121 of *H. influenzae* (accession number U32801)

ORF133 and HI121 show 57% aa identity in 363aa overlap:

	Orf133:	31	IYEPVLKYYKGRANNHNSVISIADFQDYFMFASYSRTHRMPIQEMYFSQIGDSGVHTA	90
40	Hi121:	563	I EEP+L K G K+A NHS ++SA+ DYMFEM+YSRTHRMENIQEM+FSQ+ ++GV+TA	622
	Orf133:	91	LKPERANTWQFGEXTYKKGLLKQDDTLGLKLVGYRSRIDNYIHNVYKGVWLNDGIPSWF	150
	Hi121:	623	LKPE+T+T+Q GF TYKKG L QDD LG+KLVGYRS I NYIHNVG WW ++W+	
45	Orf133:	151	SSTGLAYTIQHRFXDKVHXXXXXXXXYDYGRFTTNLSYAYQKSTQPTNFSDASESFNN	210
	Hi121:	681	ESNGFKYTIHQHNYKPIVKKSGVELEINYDMGRFFANVSYAYQRTNQPTNYADASPRFNN	740
50	Orf133:	211	ASKEDQLKQGYGLSRVSLPRDYGRLEVGTWRWGNKLTGLGAMRYFGKSIRATAEERYID	270
	Hi121:	741	ASQEDTLKQGYGLSRVSMLEPKDYGRLELGTWRFWDQKLTGLAARYGKSKRATIEEYIN	800
55	Orf133:	271	GTNGGNTSNFRQLGRKRSIKQTETLARQPLIXDFNAAYEPKKNLIFRAEVKMLFDRRYIDP	330
	Hi121:	801	GSR-FKKNTLRRENYAYVKKTEIKKQPIILDHVSYPEIKDLIAEAVQNLLDKRVYDF	859
	Orf133:	331	LDAGNDAAXERYYSFDPKDKDQVTCNADKTLGNGYKGTSSKSVLTNFARGRTPLMTMS	390
60	Hi121:	860	LDAGNDAASQRYYS L-----NNSIECAQDSAC-----GSGSKTVLYNFARGRTYILSLN	910

Orf133: 391 YKF 393
YKF
HI121: 911 YKF 913

5 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF133 shows 90.8% identity over a 392aa overlap with an ORF (ORF133a) from strain A of *N.*

meningitidis:

```

10      orf133.pep                                10      20      30
                                                PGYYSDDDEFKRAFGENSPTXKKHCNRS CGI
      orf133a      FYFDAALKKDIYRLNYSNTNTVGVRFGGKYTGYYXSDDEFKRAFGENSPTXKKHCNQS CGI
                  450      460      470      480      490      500

15      orf133.pep                                40      50      60      70      80      90
      YEPVLKKYGGKRRANNHSVISISADFGDYFMPFASYSRTHRMENIQEMYPFSGIGDSGVHTAL
      orf133a      YEPVLKKYGGKRRANNHSVISISADFGDYFMPFASYSRTHRMENIQEMYPFSGIGDSGVHTAL
                  510      520      530      540      550      560

20      orf133.pep                                100      110      120      130      140      150
      KPERANTWQFGFXTYKGLLKQDDTLGLKLIVGYRSRIDNYIHNVYGKWDNLNGDIPSPWSV
      orf133a      KPERANTWQFGFNTYKGLLKQDDTLGLKLIVGYRSRIDNYIHNVYGKWDNLNGNIPSPWSV
                  570      580      590      600      610      620

25      orf133.pep                                160      170      180      190      200      210
      STGLAYTIQHRXFXDKVHQXXXXXXXXXYDYGRFFTNLSYAYQKSTQPTNFS DASESFNNA
      orf133a      STGLAYTIQHRNFKDKVHKGFLELNYYDYGRFFTNLSYAYQKSTQPTNFS DASESFNNA
                  630      640      650      660      670      680

30      orf133.pep                                220      230      240      250      260      270
      SKEDQLKQGYGLSRVSALPRDYGRLEVGTRWLGNKLTGLGAMRYFGKSI RATAEERYIDG
      orf133a      SKEDQLKQGYGLSRVSALPRDYGRLEVGTRWLGNKLTGLGAMRYFGKSI RATAEERYIDX
                  690      700      710      720      730      740

35      orf133.pep                                280      290      300      310      320      330
      TNGGNTSNFRQLGKRISIKQTETLARQLIXDFNAAYEPKKNLIFRAEVKNLFD RRYIDEL
      orf133a      TNGXXTSNFRQLGKRISIKQTETLARQLIFDXAAYEPKKNLIFRAEVKNLFD RRYIDEL
                  750      760      770      780      790      800

40      orf133.pep                                340      350      360      370      380      390
      DAGNDAAXERYYSFDPKDKDXDVTGNADKTLGNGYGGTGSKSVLTNFARGRTFLMTMSY
      orf133a      DAGNDAARTQRYYSFDPKDKDEEVTGNDNTLGNNGYGGTGSKSVLTNFARGRTFLMTMSY
                  810      820      830      840      850      860

50      orf133.pep      KFX
      orf133a      KFX
                  870

```

55 A partial ORF133a nucleotide sequence <SEQ ID 879> is:

```

1  AAAGACAAAA  AAGTGTTCAC  CGATGCGCGT  GCGGTATCGA  CCCGTGAGGA
51  TATATATCAA  TCCANGCGAA  ACCTCGACAA  CATCGTACGC  ANCATCCCCG
101  GTGCGTTTAC  ACANCAANAT  AAAAGTCGAG  CGNTTGTGTC  TTGAATATT
151  CGGCGCGACA  CGGGGTTCGG  CGGGGTCAAT  ACNATGGTNG  AGCGATCAC
201  NCANACCTTT  TATTCGACTT  CTACCGATGC  GGGCAGGGCA  GGGGTTCAT
251  CTCAATTCGG  TGCACTCTGC  GACAGCAATT  TTAATGCCGG  ACTGATGTC
301  GTCAAAGGCA  GCTTCAGCGG  CTCGGCAGCG  ATCAACAGCC  TTGCCGGTTC
351  GCGCAATCTG  CGGACTTTAN  GCGTGGATGA  TGTCGTTTCA  GGCAATANTA
401  CNTACGGCCT  GCTGCTAAAA  GGTCTGACCG  GCACCAATTC  AACCAAAGGT

```

451 AATGCCGATGG CGGCGATAGG TGGCGGCAAA TGGCTGGAAA GCGGAGCATC
 501 TGTGCGGTGTG CTTTGGCGGC ACAGGTCGCG CAGGCTTACC CAAATATTACC
 551 GCGTGGGCGG CGGCGGCGAG CACATCGGAA ATTTTGGCGG GGAATATCTCG
 601 GACGACGCGA AGCAACGATA TTTTGGACAA GAAGGCGGGT TGAATTTCAA
 651 TTCCACACAGC GSAAAATGGG AGCGGGATTT CCAAAAGTCG TACTGGCAAAA
 701 CCAAGTGTGTA TCAAAATATC GATGCCCCCC AAGAATTGCA AAAATACATC
 751 GAAGGTCATG ATAAAAGCTG GCGGGAAAAC CTGGCGCGCG AATACGCAT
 801 CACCCCATCT GATCCGTCCA GCCTGAAGCN GCAGTCGGCA GGCAACCTGT
 851 TTAATTTGGA ATACGACGGC GTATTCAATA AATACACGGC GCAATTTTCG
 901 GATTTTAAACA CCAAAATCGG CAGCGCGAAA ATCATCAACC GCAATTTATCA
 951 ATTCAATTAC GTTTGTCTTT TGAACCCGTA TACCAACCTC AATCTGACCG
 1001 CAGCCTACAA TTTGGGCGAG CAGAAATATC CGAAAGGCTC GAAGTTTACA
 1051 GGCTGGGGGC TTTTNAAGA TTTTGAACCC TACACAACGC CAAAATCTCT
 1101 CGACCTCANC AACACCTCCA CTTTCCGGCT GCCCGGTGAA ACCGAGTTGC
 1151 AAACCACTTT GGGCTTCAAT TATTTCCACA ACGAATACGG CAAAACCTCG
 1201 TTTCTTGAAG AATTGGGGCT GTTTTTCGAC GTTCGGGATC ANGACAACGG
 1251 GCTTTATTC TATTGGGGGC GGTTTAAGGG CGATAAAGGG CTGTGCGCCC
 1301 AAAAATCAAC CATTTGTCGA CCGGCGCGCA GCGCAATATT CACACAGCTC
 1351 TACTTCGATG CGCGCTCRAA AAAAGACATT TACCGCTTAA ACTACAGGTC
 1401 CAATCCCTTC GCGTACCTTT TCGCGGCGMA ATATACGGGC TATACNGCT
 1451 CGGATGACGA ATTTAAGGCG GCAATCGGAG AAACTCGCC GACATACANG
 1501 AAACATTGCA ACCAGAGCTG CGGAATTTAT GAACCGGTAT TGAANAATAA
 1551 CGGCAAAAGG CGCGCCAACA ACCATTTCGT CAGCATTAGT GCGGACTTCG
 1601 GCGATTATTT CATGCCGTTT GCCGCTATT CGCGCACACA CCGTATGCC
 1651 AACATCCAAG AAATGTATTT TTCCCAATCT GCGGACTCCG CGGTTACAC
 1701 CGCCTTAAAA CCAGAGCGCG CAACACCTTG CCAATTTGGC TTCAATACCT
 1751 ATAAAAAAGG ATTGTTAAAA CAAGATGATA TATTAGGATT AAACTGGTC
 1801 GGCTACCGCA GCCGCATCGA CNACTACATC CACAACGTTT ACGGGAATG
 1851 GTGGGATTGT AACGGGAATA TTCCGAGCTG GGTCAACGAG ACCGGGCTTG
 1901 CCTACACCAT CCAACACCGC AATTTCAAAG ACAAAGTGCA CAAACACGGT
 1951 TTTGAGTTGG AGCTGAATTA CGATTATNNG CTTTTTTCAC CCAACCTTTC
 2001 TTACGCGCTAT CAAAAAAGCA CGCAACCGAC CAACCTTCAGC GATGCGAGCG
 2051 AATCGCGCAA CAATGCGTCC AAAGAAGACC AACTCAACAA AGGTTATGNG
 2101 TTGAGCAGGG TTGCGAGATG CGCGGAGATG TGGAGACTCG TGGAGACTCG
 2151 TACGCGCTGG TTGCGCAACA AACTGACTTT GCGCGCGGCG ATGCGCTATT
 2201 TCGGCAAGAG CATCCGCGCG ACGGCTGAAG AACGCTATAT CGACNGCAC
 2251 AATGGGNGAN NTACCGCAA TTTCCGCAA CTGGGCAAGC GTTCCATCAN
 2301 ACAAACCGAA ACCCTTGCCC GCCAGCCTTT GATTTTGTAT TTNATCGCG
 2351 CTTACGAGCC GAAGAAAAAN CTTATTTCCT CGCGCGAAGT CAAAATCTG
 2401 TTGACGAGCG GTTATATCGA TCCGCTCGAT CGCGGCAATG ATGCGGCAAC
 2451 CGAGCGTTAT TACAGTTCTG TCGACCCGAA AGACAAGGAC GAAGAAGTAA
 2501 CGTGTAAATG TGAATAACAG TTATGCAACG CCAATAACGG CGGCACAAGC
 2551 AAAAGCGTAT TGACCAATTT TGCACGCGGA CNACTTTT TGAATACGAT
 2601 GAGCTACAG TTTTAA

This encodes a protein having (partial) amino acid sequence <SEQ ID 880>:

1 KDKKVFTDAR AVSTRQDIFK SXENLDNIVR XIPGFTXQK KSSGXVSLNI
 51 RXDSGFRGVN TMDVGITKTF YSTSTDAGRA GSSGQFASV DSNFXAGLDV
 101 VKGSFSGSAG INSLAGSANL RTLXVDVVQV GNXYGLLLK GLTGTNSTRG
 151 NMAAIAIGARK WLESGASVGV LYHSRRSVA QNRYVGGGG GLTNFGAEYL
 201 ERKQVFEFO EGGILSNMS GSKWERQKRS YWKTKWYQKY DAPLEKQYI
 251 EBGHDKSWREN LAPQYDITPI DPSRLKKQSA GNMLELYDYG VBNKTLQER
 301 DLNTKIGSRK IINRNYQPNY GLSLNYPYTNL NITAAVNSGR QKYPKGSKEP
 351 GWGLXKDFET YNNAKILDLX NTSTFRLPRE TELQTLGPN YFHEYQKNR
 401 FFEELGLFFD GPDNDNGLYS YLGRFKGDKG LLPKSTIVQ PAGSOYPNF
 451 YFDAALKKDI YRLNYSNTNV GYRFGGXYTG YXSDDEFEKR AFGENSPTYX
 501 KHCNQSCGIIY EPVLKKYGKK RANNHVSIS ADGDFYFMF ASYSRTHRMP
 551 NIQEMYFSQI GDSGVHTALK PERANTWQFG FNTYKGLLLK QDDILGLKLV
 601 GYRSTRIDYI HNIVYGKWWL NGNIPSWVSS TGLAYTIQHR NFKDKVHKHG
 651 FELELNYDYX RFFTNLSYAY QKSTQPTNFS DASESPNNAS KEDQLKQGY
 701 LSRVSALPRD YGRLEVGTWR LGNKLTLGMA MRYFPGKSIRA TAEERYIDXT
 751 NGXKTSNFRQ LGRKRSXQTE TLAQPLIFD XYAYEPKKX LIFRAEVKNL
 801 FDRRYIDPLD AGNDAATQRY YSSFDPKDKD EEVTCNDNDT LCNKGKYGTS
 851 KSVLTNFARG XTFLITMSYK F*

ORF133a and ORF133-1 show 94.3% identity in 871 aa overlap:

orf133a.pep 10 20 30 40
 KDKKVFTDARAVSTRQDIFKSXENLDNIVRXIPGFTXQXKS

	orf133-1	EAQIQVLEDVHVKAKRVPKDKKVFTDARAVSTRQDIPKSSSENLDNIVRSIPGAFQTQDDKS	10	20	30	40	50	60
5			50	60	70	80	90	100
	orf133a.pep	SGXVSLNIRKDSGGRVNTMVDGITXTFYSTSTAGRAGGSSQFGASVDSNFXAGLDVVK						
	orf133-1	SGIVSLNIRGDSGGRVNTMVDGITQTFTFYSTSTAGRAGGSSQFGASVDSNFIAGLDVVK	70	80	90	100	110	120
10			110	120	130	140	150	160
	orf133a.pep	GSFSGSAGINSLAGSANLRTLXVDDVVGQNXITGLLLKGLTGTNSTKGNAMAAIGARKWL						
	orf133-1	GSFSGSAGINSLAGSANLRTLGVDDVVGQNNITGLLLKGLTGTNSTKGNAMAAIGARKWL	130	140	150	160	170	180
15			170	180	190	200	210	220
	orf133a.pep	ESGASVGLYGHSSRRSVAQNYRVGGGGHICNFGAYLERRKQRYFEQEGGLKFNSNSGK						
20			190	200	210	220	230	240
	orf133-1	ESGASVGLYGHSSRRSVAQNYRVGGGGHICNFGAYLERRKQRYFVQEGALKFNSDSGK						
25			230	240	250	260	270	280
	orf133a.pep	WERDFQKSYWKTKWYQKYDAPQELQKYIEGHDKSWRENLAPOYDITPIDPSSLKXQSAGN						
	orf133-1	WERDLQRQQWKYKPKYKNYNN-QELQKYIEHDKSWRENLXPOYDITPIDPSSLKXQSAGN	250	260	270	280	290	
30			290	300	310	320	330	340
	orf133a.pep	LFKLEYDGVFNKYTAQFRDLNLTIGSRKIINRNYQFNYGLSLNPYTNLNLTAAYNSGRQK						
	orf133-1	LFKLEYDGVFNKYTAQFRDLNLTIGSRKIINRNYQFNYGLSLNPYTNLNLTAAYNSGRQK	300	310	320	330	340	350
35			350	360	370	380	390	400
	orf133a.pep	YPKGSKFTGWGLXKDFETYNNAKILDLNLTSTFRLPRETELQTTLGPNYFHNEYGKNRFP						
	orf133-1	YPKGSKFTGWGLXKDFETYNNAKILDLNLTATFRLPRETELQTTLGPNYFHNEYGKNRFP	360	370	380	390	400	410
40			410	420	430	440	450	460
	orf133a.pep	EELGLFFDGPDXDNGLYSYLGRFKGDKGLLPQKSTIVQPAGSQYFNTFYDAALKKDIYR						
	orf133-1	EELGLFFDGPDXDNGLYSYLGRFKGDKGLLPQKSTIVQPAGSQYFNTFYDAALKKDIYR	420	430	440	450	460	470
45			470	480	490	500	510	520
	orf133a.pep	LNYSNTNVGYRFGGXITGYYXSDDEFKRAFGENSPTYXKHCHNQSGIYFVLKKYGGKRA						
50			480	490	500	510	520	530
	orf133-1	LNYSNTNVGYRFGGEYTGYYXSDDEFKRAFGENSPTYXKHCHNQSGIYFVLKKYGGKRA						
55			530	540	550	560	570	580
	orf133a.pep	NNHSVSISADFGDYFMPFASYSRTHRMFNIQEMYFSQIGDSGVHTALKPERANTWQGFEN						
	orf133-1	NNHSVSISADFGDYFMPFASYSRTHRMFNIQEMYFSQIGDSGVHTALKPERANTWQGFEN	540	550	560	570	580	590
60			590	600	610	620	630	640
	orf133a.pep	TYKGLLKQDDILGLKLVGYRSRIDXYIHNVYKGWDLNGNIPSWVSTGLAYTIQHRNF						
	orf133-1	TYKGLLKQDDTLGLKLVGYRSRIDNYIHNVYKGWDLNGDIPSWVSTGLAYTIQHRNF	600	610	620	630	640	650
65			650	660	670	680	690	700
	orf133a.pep	KDKVHKHGFEELELNLDYXRFFTNLSYAYQKSTQPTNFSDAESPNNASKEDQLKQGYGLS						
	orf133-1	KDKVHKHGFEELELNLDYGRFFTNLSYAYQKSTQPTNFSDAESPNNASKEDQLKQGYGLS	660	670	680	690	700	710

		710	720	730	740	750	760
	orf133a.pep	RVSAALPRDYGRLEVGTNRWLGNKLTGGAMRYFGKSIRATAEERYIDXTNGXXTSNFRQLG					
5	orf133-1	RVSAALPRDYGRLEVGTNRWLGNKLTGGAMRYFGKSIRATAEERYIDXTNGXXTSNFRQLG					
		720	730	740	750	760	770
	orf133a.pep	770	780	790	800	810	820
10	orf133-1	KRSIKQTETLARQPLIFDXYAAEYKPKKLI FRAEVKNLFDRRYIDPLDAGNDAATQRYYS					
		780	790	800	810	820	830
	orf133a.pep	830	840	850	860	870	
15	orf133-1	SFDPKDKDEEVTCDNDNTLCNGKYGGTSKSVLTNFARGXTFLITMSYKFX					
		840	850	860	870	880	

Homology with a predicted ORF from *N.gonorrhoeae*

- 20 ORF133 shows 92.3% identity over 392 aa overlap with a predicted ORF (ORF133ng) from *N. gonorrhoeae*:

	orf133.pep	PGYGSDDDEFKRAFGENSPTXKKHCNRSCTGI	31
25	orf133ng	FYFDAALKKDIYRLNYSTNAINYRFGGEYTGYYGSGENEKFRAGFENPAYKEHCDPSCGL	560
	orf133.pep	YEPVLKKYKKRRANNHVSISADFGDYFMPFASYSRTHRMPNIQEMYFSQIGDSGVHTAL	91
	orf133ng	YEPVLKKYKKRRANNHVSISADFGDYFMPFAGYSRTHRMPNIQEMYFSQIGDSGVHTAL	620
30	orf133.pep	KPERANTWQFGEXTYKKGLLKQDDTLGLKLVGYRSRIDNYIHMVYQKWDLNGDIPSWVS	151
	orf133ng	KPERANTWQFGENTYKKGLLKQDDTLGLKLVGYRSRIDNYIHMVYQKWDLNGDIPSWVG	680
35	orf133.pep	STGLAYTIQHRFXDKVHQXXXXXXXYDYGRFFTNLSYAYQKSTQPTNFSDAESPNNNA	211
	orf133ng	STGLAYTIHRHNFKDKVHKHGFLELNYDYGRFFTNLSYAYQKSTQPTNFSDAESPNNNA	740
40	orf133.pep	SKEDQLKQGYGLSRVSALPRDYGRLEVGTNRWLGNKLTGGAMRYFGKSIRATAEERYIDG	271
	orf133ng	SKEDQLKQGYGLSRVSALPRDYGRLEVGTNRWLGNKLTGGAMRYFGKSIRATAEERYIDG	800
	orf133.pep	TNGGNTSNFRQLGKRSIKQTETLARQPLIXDFNAAYEYKPKKLI FRAEVKNLFDRRYIDPL	331
45	orf133ng	TNGGNTSNVRQLGKRSIKQTETLARQPLIFDXYAAEYKPKKLI FRAEVKNLFDRRYIDPL	860
	orf133.pep	DAGNDAAXERYYSFDPKDKDXDVTCDNADKTLGNGKYGGTSKSVLTNFARGRTFLMTMSY	391
	orf133ng	DAGNDAATQRYYSFDPKDKDEEDVTCDNADKTLGNGKYGGTSKSVLTNFARGRTFLMTMSY	920
50	orf133.pep	KF 393	
	orf133ng	KF 922	

The complete length ORF133ng nucleotide sequence <SEQ ID 881> is predicted to encode a protein having amino acid sequence <SEQ ID 882>:

55	1	MRSSFLRKPI	CFYLMGVMLY	HHSYAEDAGR	AGSEAQIQVL	EDVHVKKAKRV
	51	PKDKKVFTDA	RAVSTRQDVF	KSGENLDNIV	RSIPGAFTQQ	DKSSGIVSIN
	101	IRGDSGFGRV	NTMVDGITQT	FYSTSTDAGR	AGGSSQFGAS	VDSNFIAGLD
	151	VVKGSFSGSA	GINSLAGSAN	LATLGVDVVV	QGNNTYGLLL	KGLTGTNSZK
60	201	GNAMAAICAR	KWLESASVG	VLYGHSRRGV	AGNYRVGGGG	QHIGNGFEZY
	251	LEARKQYTV	QEGGLCPNAG	SGKWEHLQR	QWTKTWYKK	YEDFQELQRY
	301	IEEHDKSWRE	NLAQYDITE	IDPSGLKQGS	AGNLLALEYD	GVEKNYTAQF
	351	RDLNTRIGSR	KIINRYQFN	YGLSLNFTTN	INLRYAYNSG	RQKYPKGAKF
	401	TGWGLLKDFE	TYNNAKIDL	NTATFRLPR	ETELQTTLGF	NFFNNEYGRN

5	451	RFPEELGLFF	DGPDQNGLY	SYLGRFKGDK	GILLPOKSTIV	QPAGSQYFNT
	501	PYFDAALKKD	IYRLNYSNA	INRYFGGEYT	GYGSENEFK	RAFGENSPAY
	551	KEHCDPSCGL	YEPVLKYKYG	KRANNHSVSI	SADPGDYFIR	FAGYSRTHRM
	601	PNIQEMYFSQ	IGDSGVHTAL	KPERANTWOF	GFNTYKKGLL	KODDILGLKL
	651	VGYRSRIDNY	IHNVYKXWD	LNQDIPSWVG	STGLAYTIRH	RNFKDKVHKH
	701	GFELEINNDY	GRFFTNLSYA	YOKSTOPTNF	SDASESPFNA	SKEDQLKQGY
	751	GLSRVSALEP	DYGRLEVETR	WLNKLTGG	AMRYFGKSIR	ATAERYIDG
	801	TNGGNTSNVR	QLGKRSIKQT	ETLARQPLIF	DFYAAYEPKK	NLIFRAEVKN
	851	LFDRYRIDPL	DAGNDAATQR	YSSFDPKDK	DEDVTCNADK	TLCNGKYGGT
10	901	SKSVLTNFR	GRTFLMTMSY	KF*		

A variant was also identified, being encoded by the gonococcal DNA sequence <SEQ ID 883>:

15	1	ATGAGATCTT	CTTCCGGT	GAAGCCGATT	TGTTTTATC	TTATGGGTGT
	51	TATGCTATAT	CATCATAGTT	ATGCCGAAGA	TGCAGGGCCG	CGGGSCAGCG
	101	AGGCGCAGAT	ACAGGTTTTG	GAAATGTGCG	ACGTCAAGCG	GAAGCCCGTA
	151	CCGAAGACAA	AAAAAGTGTT	TACCGATGCG	CGTCCGCTAT	GACCCGCTca
	201	gGATGTGTT	AAATCCGGCG	AAAACTCGA	CAACATCGTA	CGCAGCATA
	251	CCGGTCCGTT	TACACAGCAA	GATAAAGCT	CGGCAATTTG	GTCTTTGAAT
	301	ATTCGCGCGG	ACAGCGGGTT	CGGGCGGGTC	AATACGATGT	TGGACGGCAT
	351	CACGCAACAG	TTTTATACGA	CTTCTACGGA	TCGGCGGCTG	CGACGCGGAT
20	401	CATTCTCAATT	CGGTGACCTT	CTCGACAGCA	ATTTTATTGC	CGAGCTGGAT
	451	GTCTGCAAG	GCAGCTTCAG	CGGCTCGCGA	GCATCAACA	GCCTTGCCGG
	501	TTCCGGCGAAT	CTCGCGACTT	TAGCGTGGTA	TGACGTCGTT	CAGGCGCAATA
	551	ATACCTACGG	CTGCTGCTGA	AAAGGTCTGA	CCGCGACCAA	TTCAACCAAA
	601	GGTAATGCGA	TGGCGCGGAT	AGGTGCGCGC	AAATGGCTGG	AAAGCGCGAGC
25	651	GTCTGTCCGT	GTGCTTTACG	GGCACAGCAG	CGCGGGCGTG	CGCGAAAAAT
	701	ACCGCGTGGG	CGGCGGCGGG	CAGCACATCG	GAAATTTTGG	TGAGAAATAT
	751	CTGGAACGGC	GCAACACGCA	ATATTTTGTA	CAAGAGGGTG	GTTTGAAAT
	801	CAATGCCGGC	AGCGGAAAAA	GGGAACGGGA	TTTGCAAGG	CAATCATGGA
	851	AAACCAAGTG	GTATAAAAAA	TACGAAGACC	CCCAAGAACT	GCAAAAAATC
30	901	ATCGAAGAGC	ATGATAAAAA	CTGGCGGGAA	AACCTGGCGC	CGCAATACGA
	951	CATCACCCCC	ATCGATCCGT	CCGGCTCGAA	GCAGCAGTCG	CGAGGCAATC
	1001	TGTTTAAATT	GGAAATACAC	GGCGTATCCA	ATAAATACAC	CGCGCAATTT
	1051	CGGATATATA	ACACCAAGAT	CGGACGCGCG	AAAAATCATCA	ACCGCAATTA
	1101	TCATATTCAT	TACGCTTTCT	CTTTGCAACC	GTATACCAAC	CTCATCTGTA
35	1151	CCGCGAGCTA	CAATTCCGGC	AGGCAGAAAT	ATTCGAAAGG	AGGCGAGCTT
	1201	ACAGGCTGGG	GGCTTTTAAA	AGATTTTGAA	ACCTCAACAA	ACGCGAAAT
	1251	CCTCGACCTC	ACACAACACG	CCACCTTCCG	GCTGCCCGCG	GAAACCGAAT
	1301	TGCAAAACAC	TTTGGGCTTC	AATTATTTCG	ACAACGAATA	CGGCAAAAAC
40	1351	CGCTTTCCTG	AGAATTTGGG	GCTGTTTTTC	ATCGGCTCTG	ATCAGACAAA
	1401	CGGGCTTTAT	TCCTATTTCG	GGCGGTTTAA	GGGCGATAAA	GGGCTGTTCG
	1451	CTCAAAAATC	AACCATTTGC	CAACCGCGCG	CGACGCAATA	TTTCAACACG
	1501	TTCTACTTCG	ATGCGCGGCT	CAAAAAGAAC	ATTAATCCGCT	TAAACTACAG
	1551	CACCAATGCA	ATCAACTACC	GTTTCCGGCG	CGAATATACG	GGCTATTACG
45	1601	GCTCGGAAAA	CGAATTAAAG	CGGGCATTCG	GAGAAAACCT	CGCGGCATAC
	1651	RAAGAACATT	CGCAACCGAG	CTGCGGGCTT	TATGAACCCG	TATTGAAAAA
	1701	ATACGGCAAA	AAGCGGGCCA	ACAACCATTC	GGTCAGCAAT	AGTCGGGACT
	1751	TCGGCGATTA	TTTCATGCGG	TTGCGCGGCT	ATTGCGGCGT	ACAGCGGTATG
	1801	CCCAACATCC	AGAATATGTA	TTTTTCCCAA	ATTCGGGCGT	CCGGCGTCTA
50	1851	CACGCGCTTA	AAACCAAGAG	CGGCAACACG	TTTGCATATT	GGCTTCAATG
	1901	CTTATAAAAA	AGGATGTGTA	AAACAAGATG	ATTTATAGG	ATTGAACATG
	1951	GTCCGCTACC	GCAGCCGCGT	TGACAACTAC	ATCCACAACG	TTTACGGGAA
	2001	ATGTTGGGAT	TTGAACGGGG	ATATTCCGAG	CTGGGTCCGG	AGCACCGGGC
	2051	TTGCCTACAC	CATCCGACAC	CGCAATTTCA	AAGACAAGAT	GCACAAACAC
55	2101	GGTTTGTAGC	TGGAGCTGAA	TTACGATTAT	GGGGTTTTTT	TCACCAACCT
	2151	TTCTTACGCC	TATCAAAAAA	CAGCAGCAAC	GACCAATTTT	AGCGATTGGA
	2201	CGGAATCGCC	CAACAATGCC	tccaaAGAA	ACCAACTCAA	ACAAGGTTAT
	2251	GGGCTGAGCA	GGGTTTCCCG	CTCGCCGCGA	GATTACGGAC	GTTTGGAAGT
	2301	CGGTACCGCG	TGGTTGGGCA	ACAAACTGAC	TTTGGGCGCG	CGGATg-cGCT
60	2351	ATTTCCGGCA	GAGCATCCGC	GCGACGGCTG	AAGAAGCGTA	TATCGACGGC
	2401	ACCAACGGGG	GAATATACAG	CAATGTCCGC	CACCTGGGCA	AGCGTTCCAT
	2451	CAACAACACC	GAACCCCTTG	CCCGACGCGC	TTTGATTTTT	GATTTTTACG
	2501	CCGCTTACGA	GCACAGCTTA	CGATCTATTG	TCGCGCGCGA	AGTCAAAAAC
	2551	CTGTTCACAC	GGCGTTATAT	CGATCCGCTC	GATCGCGCGA	ATGATCGCGA
	2601	AACGCAAGCT	TATTACAGCT	CGTTCGACCC	GAAGACAAG	GACCAAGACG
65	2651	TAACTGTGTA	TGCTGATAAA	ACGTTGTGCA	ACGCGAAATA	CGGCGGACCA
	2701	AGCAAAAGCG	TATTGACCAA	TTTCCGACGC	GGACGACCTT	TCTTGATGAC
	2751	GATGAGCTAC	AAGTTTAA			

This corresponds to the amino acid sequence <SEQ ID 884; ORF133ng-1>:

```

1  MRSSFLKPI CFYLMGVMLY HHSYAEDAGR AGSEAQIQVL EDVHVKAKRV
5  51  PKDKKVFTDA RAVSTRQDVF KSGENLDNI VRSIPGAFTQQ DKSSGIVSLN
101 IRGDSGSGFRV NTMVDGITQT FYSTSTDA GRAGSSQFGAS VDSNFIAGLD
151 VVKGSGSGSA GINSLAGSAN LRTLGVDDVV QGNNTYGLLL KGLTGTNSTK
201 GNAMAAIGAR KWLESGASVG VLYGHSRRGV AQNYRVGGGG QHIGNFGEEY
251 LERRKQQYFV QEGGLKFNAG SGKWERDLQR QYWKTKWYKK YEDPQLQKY
301 IEEHDKSWRE NLAPQYDITP IDPSGLKQGS AGNLFKLEYD GVFNKYTAQF
351 RDLNTRIGSR KLINRNYQFN YGSLNPNYTN LNLTAAYNSG RQKYPKGAKE
10 401 TGWGLLKDFE TYNNAKILD LNTATFRFLR ETELQTTLLG NYFNHEYGNF
451 RFPEELGLFF DGPQQDNGLY SYLGRFKGDK GLLPQKSTIV QPAGSQYFTN
501 FYFDAALKKO IYRLAVSYNA INYRFGGEYT GYSGSENEFK RAFGENSPAY
551 KEHCDPSCGL YEPVLKKYKK KRANNHSVSI SADEGDFMP FAGYSRTHRM
15 601 PNIQEMYFSQ IGDGCVHTAL KPERANTWQF GENTYKGLL KQDDILGLKL
651 VGYRSRIDNY IHNVYKGWWD LNGDIPSWVG STGLAYTIRH RNFKDKVHKH
701 GFLEELNYDY GRFFTNLSYA YKSTQPTNF SDASESPNNA SKEDQLQKQY
751 GLSRVSALPR DYGRLEVGTR WLGNKLTLLG AMRYFGKSI R ATAEERYIDG
801 TNGGNTSNVR QLGKRSIKOT ETLARQLIF DFYAAEYPKK NLIHRAEVKN
851 LFDRRYIDPL DAGNDAATQR YSSSFDPKDK DEDVTCNADK TLCNGKYGGT
20 901 SKSVLTNPAR GRTEFLMTSY KF*

```

ORF133ng-1 and ORF133-1 show 96.2% identity in 889 aa overlap:

```

              10      20      30      40      50      60
orf133ng-1.pep SFRLKPICFYLMGVMLYHHSYAEDAGRAGSEAQIQVLEDVHVKAKRVPKDKKVFTDARAV
25 orf133-1                                     EAQIQVLEDVHVKAKRVPKDKKVFTDARAV
              10      30
              70      80      90      100     110     120
orf133ng-1.pep STRQDVKFSGENLDNIVRISIPGAFTQQDKSSGIVSLNIRGDSGSGFRVNTMVDGITQTFYS
30 orf133-1     STRQDIFKSGENLDNIVRISIPGAFTQQDKSSGIVSLNIRGDSGSGFRVNTMVDGITQTFYS
              40      50      60      70      80      90
              130     140     150     160     170     180
orf133ng-1.pep TSTDAGRAGGSSQFGASVDSNFIAGLDVVKGSGSGAGINSLAGSANLRTLGVDDVVQGN
35 orf133-1     TSTDAGRAGGSSQFGASVDSNFIAGLDVVKGSGSGAGINSLAGSANLRTLGVDDVVQGN
              100     110     120     130     140     150
              190     200     210     220     230     240
orf133ng-1.pep NTYGLLLKGLTGTNSTKGNAMAAIGARKWLESGASVGVLVYGHSSRGVAQNYRVGGGGQHI
40 orf133-1     NTYGLLLKGLTGTNSTKGNAMAAIGARKWLESGASVGVLVYGHSSRGVAQNYRVGGGGQHI
              160     170     180     190     200     210
              250     260     270     280     290     300
orf133ng-1.pep GNFGEELERRKQQYFVQEGGLKFNAGSGKWERDLQRQYWKTKWYKKYEDPQLQKYIEE
45 orf133-1     GNFGAELERRKQRYFVQEGALKFNSDSGKWERDLQRQYWKYKPYKNYN-NELQKYIEE
              220     230     240     250     260
              310     320     330     340     350     360
orf133ng-1.pep HDKSWRENLAPOYDITPIDPSGLKQGSAGNLFKLEYDGVFNKYTAQFRLNTRIGSRKII
55 orf133-1     HDKSWRENLAPOYDITPIDPSSLKQGSAGNLFKLEYDGVFNKYTAQFRLNTRIGSRKII
              270     280     290     310     320
              370     380     390     400     410     420
orf133ng-1.pep NRNYQFNGLSLNPNYTNLNLTAAYNSGRQYKPKGAKFTGWGLLKDFEYTNNAKILDNLNT
60 orf133-1     NRNYQFNGLSLNPNYTNLNLTAAYNSGRQYKPKGSKFTGWGLLKDFEYTNNAKILDNLNT
              330     340     350     360     370     380
              430     440     450     460     470     480
orf133ng-1.pep ATRFLPRETELQTLGFNYFHNEYGNRFPPEELGLFFDGPQQDNGLYSYLGRFKGDKGGLL
65

```


	orf133-1	ATFRLPRETELQTTILGFNYFHNEYGKNRFPEELGLFFDGDQDNGLSYLGRFKGDKGLL
		390 400 410 420 430 440
5	orf133ng-1.pep	490 500 510 520 530 540
	orf133-1	PQKSTIVQAPAGSQYFNTFYFDAALKKDIYRLNYSNTNAINYRFGGEYTGYYGSENEFKRAF
		450 460 470 480 490 500
10	orf133ng-1.pep	550 560 570 580 590 600
	orf133-1	GENSPAYKEHCDPSCGLYEPVLKKYGGKRRANNHSVSISADFGDYFMPFAGYSRTHRMPNI
		510 520 530 540 550 560
15	orf133ng-1.pep	610 620 630 640 650 660
	orf133-1	QEMYSQIGDSGVHTALKEPERANTWQFGFNTYKGLLKQDDILGLKLVGYRSRIDNYIHN
20	orf133ng-1.pep	670 680 690 700 710 720
	orf133-1	VEYKQWDLNGDIPSWVGSTGLAYTIHRNFKDKVKHGFLELNIDYGRFFNTLSYAYQK
25	orf133ng-1.pep	730 740 750 760 770 780
	orf133-1	STQPTNFSDAESPNNASKEDQLKQGYGLSRVSALPRDYGRLEVGRTRWLNKLTGGAMR
30	orf133ng-1.pep	790 800 810 820 830 840
	orf133-1	YFGKSIIRATAEERYIDGTNGGNTSNVRQLGKRSIKQTETLARQLIFDFYAAEYEPKKNLI
35	orf133ng-1.pep	850 860 870 880 890 900
	orf133-1	FRAEVKNLFDRIYIDPLDAGNDAATQRYYSFDPKDKDEDVTCNADKTLGNGKYGGTSKS
40	orf133ng-1.pep	910 920
	orf133-1	VLTNFARGRTFLMTMSYKFX
45	orf133ng-1.pep	870 880
	orf133-1	VLTNFARGRTFLMTMSYKFX

In addition, ORF133ng-1 is homologous to a TonB-dependent receptor in *H. influenzae*:

	sp P45114 YC17_HABIN PROBABLE TONB-DEPENDENT RECEPTOR Hii1217 PRECURSOR
	>gi 1075372 pf G64110 transferrin binding protein 1 precursor (tblp1) homolog -
55	Haemophilus influenzae (strain Rd KW20) >gi 1574147 (U32801) transferrin binding
	protein 1 precursor (tblp1) [Haemophilus influenzae] Length = 913
	Score = 930 bits (2377), Expect = 0.0
	Identities = 476/921 (51%), Positives = 619/921 (66%), Gaps = 72/921 (7%)
60	Query: 38 QVLEDVHVAKRVPKDKKVFTDARAVSTRQDVFVKSSENLDNIVRSIPGAFQDQKSSGIV 97
	+ L + V K + DKK FT+A+A STR++VFK + +D ++RSIPGAFQDQK SG+V
	Sbjct: 29 ETLGQIDVVEKVISNDKKPTEAKAKSTRENVFKETQTIDQVIRSIPGAFQDQKSGSVV 88
65	Query: 98 SLNIRGDSGFGVRNTMVDGITQTFYSTSTDAGRAGSSSQFGASVDNSFIAGLDVVGKGSF 157
	+NIRG++G GRVNTMVDG+TQTFYST+ D+G++GGSSSQFGA++D NFIAG+DV K +ES
	Sbjct: 89 SVNIRGENGLGRVNTMVDGVTQTFYSTALDSGQSGSSSQFGAIDPNFIAGVDVYKSNFS 148
	Query: 158 GSAGINSLAGSANRLTGLVDDVVQXXXXXXXXXXXXXXXXXAMAI GARKWLESGA 217
	G++GIN+LGSAN RTLGV+DV+ M RKLW++G

Sbjct: 149 GASGINALAGSANFRITLGVNDVITDDKPFGIILKGMGTGSNATKSNFMTMAAGRKWLNDGG 208
 Query: 218 SVCVLVYGHSSRRGVAQNYRVGGGGQHIGNFGEEYLERRKKQYFVQEGGLKFNAGSGKWED 277
 VGV+YG+S+R V+Q+YR+ GGG+ + + G+L + K+ YF + G N G+W D
 Sbjct: 209 YGVVYGYSGREVSQDYRI-GGGERLASLGQDILAKEKEAYF-RNAGYILNP-EGQWTFD 265
 Query: 278 LQRQYWK-----TKWY-----KKYEDPQELQK---YIEE 303
 L +++W +Y KK +D ++LQK IEE
 Sbjct: 266 LSKKHWSCNKPDPYQKNGDCSYRIGSAARTREILQELLTNGKKPKPIELKQKNGDIEE 325
 Query: 304 HDKSWRENLAQYDITPIDPSGLKQOSAGNLFKLEYDGVFNKYTAQFRDLNTRIGSRKII 363
 DKS+ N QY + PI+P L+ +S +L K EY AQ R L+ +IGSRKI
 Sbjct: 326 TDKSFEEN-KDQYSVAPIEPGSLQSRSRSHLLKFEYGDHQNGLGAQLRTLNDKNGISRKIE 384
 Query: 364 NRNYQFNYGLSLNPNYTNLNLTAAYNSGRQKYPKGAKFTGWGLLKDFETYNNAKILDLNNT 423
 NRNYQ NY + N Y +LNL AA+N G+ YFKG F GW + T N A I+D++N+
 Sbjct: 385 NRNYQVNYFNNNYSYLDNLMAAHNIGKTYTYPKGGFFAGWQVADKLITKNVANVDINNS 444
 Query: 424 ATRFLPRTELQTTLGFNYFHNEYGKNRFPPEELGLFFDGPQDNGLYSY--LGRFKGDKG 481
 TF LP+E +L+TTLGFNYF NEY KNRFPPEEL LF++ D GLYS+ GR+ G K
 Sbjct: 445 HTFLLPKEIDLKTLTGLFNFTNEYSKNRPPEELSLFYNDASHDQGLYSHSKRGYSGTKS 504
 Query: 482 LLEPKSTIVQAGSQYFNTFYFDAALKKDIYRLNYSNTAINRYRFGGEYTYGYYGENEFKR 541
 LLEQ+S I+QP+G Q F T YFD AL K IY LNYN + Y F GEY GY
 Sbjct: 505 LLEPQSVIILQPSGGKFKFTYVYFDALSKGIYHLNYSVNFTHYAFNGEYVGY----- 555
 Query: 542 AFGENSPAYKEHCDPSCGLYEPVLKKYKKRANNSHVSISADFGDYFMPFAGYSRTHRMP 601
 EN+ + + EP+L K G K+A NHS ++SA+ DYFMPF YSRTHRMP
 Sbjct: 556 ---ENTAGQQ-----INEPILHKS GHKKA FNHSATLSAELS DYFMPFFTYSRTHRMP 604
 Query: 602 NIQEMYFSQIGDSGVHTALKEPERANTWQGFNTYKKGGLKQDDILGLKLVGYRSRIDNYI 661
 NIQEM+FSQ+ ++GV+TALKEP+++T+Q GFNTYKKGGL QDD+LG+KLWGYRS I NYI
 Sbjct: 605 NIQEMFFSQVSNAGVNTALKEPQSDTYQLGFMNTYKKGFLTDQDVLGVKLVGYRSFKINKYI 664
 Query: 662 HNRYGKWWDLNGDIPSWVGSTGLAYTIRHRNFKDKVHKHGFELNLYDGRFFTNLSYAY 721
 HNRYG WW +P+W S G YTI H+N+K V K G ELE+NYD GRFF N+SYAY
 Sbjct: 665 HNRYGVWV---RDGMPTWAESNGFKYTIHQNKYPIVKKSGVELEINYDMGRFFANVSAY 722
 Query: 722 QKSTQPTNFDASESPNNASKEDELKQGYGLSRVSLPRDYGRLEVGTNRWLGNKLTGGA 781
 Q++ QPTN++DAS PNNAS+ED LKQGYGLSRVS LP+DYGRLE+GTRW KLT/LG A
 Sbjct: 723 QRTNQPTNYADASPRPNNASQEDILKQGYGLSRVSMPLPKDYGRLEGTNRWFDQKLT/LGIA 782
 Query: 782 MRYFGKSI RATAEERYIDGTNGGNTSNVRQLGKRSIKQTTETLARQPLIFDYAAEYFPKN 841
 RY+GKS RAT EE YI+G+ + +R+ ++K+TE + +QP+I D + +YEP K+
 Sbjct: 783 ARYFGKSKRATIEEYINGSR-FKNTLRRENYAVKRTEDIKKQPIILDLHVSYPEIKD 841
 Query: 842 LIFRAEVKNLFDRIYIDPLDAGNDAAATCRYSSFPDKDKDEVDTCNADKTLNKGYYGTS 901
 LI +AEV+NL D+RY+DPLDAGNDAA+CRYSS + + C D + C GG+
 Sbjct: 842 LIITAEVQNLLDKRYVDPLDAGNDAAASQRYSSSL-----NNSIECAQDSSAC---GGSD 892
 Query: 902 KSVLTNFARGRTFLMTMSYKF 922
 K+VL NFARGRT+++++YKF
 Sbjct: 893 KTVLYNFARGRTYILSLNYKF 913

The underlined motif in the gonococcal protein (also present in the meningococcal protein) is
 predicted to be an ATP/GTP-binding site motif A (P-loop), and the analysis suggests that these
 proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for
 vaccines or diagnostics, or for raising antibodies.

Example 104

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 885>

1 ATGAACCTGA TTTACAGTTA CATCATCGGT CAAATGGCGG TTATGGCGGT

51 TTACGCGCTC CTTGCTTCC TCGCTTTGTA CAGCTTTTTT GAAATCCTGT
 101 ACGAAACCGG CAACCTCGGC AAAGGCGAGT ACGGCATATG GGAATCCTGT
 151 GCTTACACCG CCTCAAAAT GCCGCGCGC GCTTACGAAC TGATTCOCCT
 201 CGCCGCTCCT ATCGGCGGAC TGGTCTCCT CAGCCAGCTT CGCGCGCGCA
 5 251 GCGAACTGAC CGTCATCAA GCCAGCGGCA TGAGCACCAA AAAGCTGCTG
 301 TTGATTCTGT CGCAGTTCGG TTTTATTTTT GCTATTGCCA CCGTCGCGCT
 351 CGGCGAATGG GTTGGCGCCA CACTGAGCCA AAAAGCCGAA AACATCAAAG
 401 CCGCGCCAT CAACGGCAAA ATCAGCACCG GCAATACCGG CCTTTGGCTG
 451 AAGAAAAAA ACAGCGTGAT CAATGTGCGC GAAATGTTCG CCGACCAT..

10 This corresponds to the amino acid sequence <SEQ ID 886; ORF112>:

1 MNLISRYIIR QMAVMAYAL LAFLALYSFF EILYETGNLG KSGYGIWEML
 51 GYTALKMPAR AYELIPLAVL IGGVLVLSQL AAGSELTVIK ASGMSTKLL
 101 LILSQFGFIF AIATVALGEW VAPTLSQKAE NIKAAAINGK ISTGNTGLWL
 151 KERNSVINVR EMLPDH...

15 Further work revealed further partial nucleotide sequence <SEQ ID 887>:

1 ATGAACCTGA TTTCAOGTTA CATCATCGT CAAATGCGCG TTATGGCGGT
 51 TTACGCGCTC CTTGCTTCC TCGCTTTGTA CAGCTTTTTT GAAATCCTGT
 101 ACGAAACCGG CAACCTCGGC AAAGGCGAGT ACGGCATATG GGAATCCTGT
 20 151 GCTTACACCG CCTCAAAAT GCCGCGCGC GCTTACGAAC TGATTCOCCT
 201 CGCCGCTCCT ATCGGCGGAC TGGTCTCCT CAGCCAGCTT CGCGCGCGCA
 251 GCGAACTGAC CGTCATCAA GCCAGCGGCA TGAGCACCAA AAAGCTGCTG
 301 TTGATTCTGT CGCAGTTCGG TTTTATTTTT GCTATTGCCA CCGTCGCGCT
 351 CGGCGAATGG GTTGGCGCCA CACTGAGCCA AAAAGCCGAA AACATCAAAG
 25 401 CCGCGCCAT CAACGGCAAA ATCAGCACCG GCAATACCGG CCTTTGGCTG
 451 AAGAAAAAA ACAGCTTAT CAATGTGCGC GAAATGTTCG CCGACCATAC
 501 GCTTTTGGGC ATCAAAATTT GGGCGCGCAA CGATAAAAC GAATTTGGCAG
 551 AGGCGAGTGA AGCGAATTC GCGCTTTTGA ACAGCGACGG CAGTTGGCAG
 601 TTGAAAAACA TCGCGCGCAG CACGCTTGGC GAAGACAAAG TCGAGGTCTC
 30 751 TATTGCGGCT GAAGAAACT GCGCGATTTT CGTCAACGCG AACCTGATGG
 801 ACGTATTGCT CGTCAAAACC GACCAATGT CCGTCGCGCA ACTGACCAAC
 751 TACATCCGCG ACCTCCAAAA CAACAGCCAA ACACCCCGAA CTCTACGCAAT
 801 CGCATGTTGG GCAGAAATGG TTTACCCGCG CGCAGCTGG GTGATGGCGG
 851 TCGTCGCGCT TGCCTTTACC CCGCAAAACA CCGCGCACGG CAATATGGCG
 901 TTAATACTCT TCGCGCGCAT CTGTCTCGA TTGCTGTTCG ACCTTGGCGG
 35 951 ACGGCTCTTT GGGTTTACCA GCCACTCGG...

This corresponds to the amino acid sequence <SEQ ID 888; ORF112-1>:

1 MNLISRYIIR QMAVMAYAL LAFLALYSFF EILYETGNLG KSGYGIWEML
 51 GYTALKMPAR AYELIPLAVL IGGVLVLSQL AAGSELTVIK ASGMSTKLL
 101 LILSQFGFIF AIATVALGEW VAPTLSQKAE NIKAAAINGK ISTGNTGLWL
 40 151 KERNSVINVR EMLPDHTLLG IKIWARNDKN ELAAVEADS AVLNSDGSQW
 201 LKNIRRTSLG EDKVEVSI AA EENWPISV KR NLMDVLLVKP DQMSVGLTII
 251 YIRHLQNSQ NTRIYAIWW RKLVPAAAW VMALVAFAET PQTRHGNMG
 301 LKLPFGICGX LLPHLAARLF GFTSQL...

Computer analysis of this amino acid sequence predicts two transmembrane domains and gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF112 shows 96.4% identity over a 166aa overlap with an ORF (ORF112a) from strain A of *N.*

meningitidis:

50 orf112.pep MNLISRYIIRQMAVMAYALLAFLALYSFFEILYETGNLGKSGYGIWEMLYGTALKMPAR
 orf112a MNLISRYIIRQMAVMAYALLAFLALYSFFEILYETGNLGKSGYGIWEMLYGTALKMPAR
 55 orf112.pep AYELIPLAVLIGGLVLSQLAAGSELTVIKASGMSTKLLLLILSQFGFIFAIATVALGEW

10

15

20

35

40

45

50

55

60

orf112-1 DQMSVGEITTYIRHLQNNQNTRIYAIANWKRKLYPAAAWMALVAFATPQTTRHNGMG
 orf112a.ppep LKKFGGICGLGLFHLGRLFFXTSOLYIPFLKALPTIAFALLAVLIRKQEKRX
 5 orf112-1 LKLFGGICXGLLFLHLAGRLFGFTSQL

Homology with a predicted ORF from *N.gonorrhoeae*

ORF112 shows 95.8% identity over 166aa overlap with a predicted ORF (ORF112ng) from *N. gonorrhoeae*:

10 orf112.ppep MNLISRYIIRQMAMVAVYALLAFALYLSFFEILYETGNLKGSGYIWEMLGYTALKMPAR 60
 orf112ng MNLISRYIIRQMAMVAVYALLAFALYLSFFEILYETGNLKGSGYIWEMLGYTALKMPAR 60
 15 orf112.ppep AYEILPLAVLIGGLVLSQLAAGSELTVIKASGMSTKKLLLSQFGFIFAIAATVALGEW 120
 orf112ng AYEILPLAVLIGGLVLSQLAAGSELTVIKASGMSTKKLLLSQFGFIFAIAAVALGEW 120
 orf112.ppep VAPTLSQKAEINIKAAAINGKISTGNTGLWLKEKNSVINVREMLPH 166
 orf112ng VAPTLSQKAEINIKAAAINGKISTGNTGLWLKEKTSIINVRLGMLPHTLLGIKIWARNDKN 180

The complete length ORF112ng nucleotide sequence <SEQ ID 891> is:

1 ATGAACCTGA TTTCACGTTA CATCATCCGC CAAATGGCGG TTATGGCGGT
 51 TTACGCGCTC CTTCGCTTCC TCGCTTTGTA CAGCTTTTTT GAAATCCTGT
 101 ACGAAACCGG CAACCTCGGC AAGGCGAGT ACAGCATATG GGAAATGCTG
 25 151 GGCTACACCG CCCTCAAAAT GCCGCGCCGC GCCTACGAAC TCATGCCCTCT
 201 CGCGCTCTCT ATCGGGGGAC TGGCTCTCTC CAGCGACGCTT GCGCGCGGCA
 251 GCGAACTGGC CGTCATCAAA GCGAGCGGCA TGAGCACCAA AAGCTGCTGT
 301 TTGATCTCTG CTACAGTCCG TTTTATTTT GCTATTCGCG CTCTCGCGGT
 351 CGCGCAATGG GTTGCACCCA CCGTGAGCCA AAGAGCCGAA AACATCAACA
 401 CCGCGCCATC TACGGGCAA ATCAGCAcgg cAATACCGG CCTTTggcgt
 30 451 AAGAGAAAAA ccACGATTAT CAATGTGcGc GGAATGTTGC CCGACCATAC
 501 GCTTTTGGGC ATCAAAATTT GGGCGCGCAA CGATAAAAAA GAATTGGCAG
 551 AGGCACTGGA AGCCGATTCC GCCCTTTTGA ACAGCGACGG CAGCTGGCAG
 35 601 TTGAAAAACA TCGCGCCGAG CATCATGGGT ACAGACAAAA TCGAAACATC
 651 cgcgcgcgcg GAAGAAACTT gCGCGATTGC CGTCAGACGC AACCTGATGG
 701 ACGTATTGCT CGTCAAGCCC GACCAAATGT CCGTCGCGGA GCTGACCCACC
 751 TACATCCGCG ACCTCCAAAA CAACAGCCAA AACACCCAAA TCTACGCCAT
 801 CGCATGGTGG CGTAAACTCG TTTACCCCGT CGCGCATGG GTCATGGCGC
 851 TCGTTGCCCT GCCTTTTACG CCGCAAAACA CGGCGACGG CAATATGGGC
 40 901 TAAAAACTCT TCGCGCGCAT CTGTCTCGGA TTGCTGTTCC ACCTTGCCTGG
 951 CAGGCTCTTC GGGTTTACCA GCCAACTCTA CGGCACCCCA CCCTTCTCTG
 1001 CCGCGGCACT GCCTACGATA GCCTTCGCTT TGCTCGCTGT TTGGCTGATA
 351 CGCAACAGG AAAAACGTTG A

This encodes a protein having amino acid sequence <SEQ ID 892>:

45 1 MNLISRYIIR QMAMVAVYAL IAFALYLSFF EILYETGNLKGSGYIWEML
 51 GYTRLMKMPAR AYEILPLAVL IGLLASLSQL AAGSELAVIK ASGMSTKKLL
 101 LLSQFGFIFAIAAVALGEW VAPTLSQKAE NIKAAAINGK ISTGNTGLWL
 151 KEKTSIINVRL GMLPDHTLLG IKIWARNDKN ELAAVEADS AVLNSDGSQW
 201 LKNIRRSIMG TDKIETSAEA EETWPIAVRR NIMDVLLVKP DQMSVGEITTY
 50 251 YIRHLQNNQS NTQIYAIANW RKLVPVAAW VMALVAFATP POTTRHNGMG
 301 LKLFGGICGL LFLHLAGRLF GFTSOLYGTP PFLAGALPTI AFALLAVLWLI
 351 RKQEK*

ORF112ng and ORF112-1 show 94.2% identity in 326 aa overlap:

55 orf112ng MNLISRYIIRQMAMVAVYALLAFALYLSFFEILYETGNLKGSGYIWEMLGYTALKMPAR 60
 orf112-1 MNLISRYIIRQMAMVAVYALLAFALYLSFFEILYETGNLKGSGYIWEMLGYTALKMPAR 60
 10 20 30 40 50 60
 70 80 90 100 110 120

	orf112ng	AYELMPLAVLIGGLASLSQLAAGSELAVIKASGMSTKKLLLSQFGFTFAIAAVALGEW
	orf112-1	AYELIPLAVLIGGLVSLSQLAAGSELTIVIKASGMSTKKLLLSQFGFTFAIATVALGEW
5		70 80 90 100 110 120
	orf112ng	130 140 150 160 170 180
	orf112-1	VAPTLSQKAENIKAAAINGKISTGNTGLWLKEKTSIINVRGMLPDHTLLGKIWARNDKN
10		130 140 150 160 170 180
	orf112ng	190 200 210 220 230 240
	orf112-1	ELAEAVEADSAVLNSDGSWQLKNIRRSIMGTDKIETSAAAEETWPIAVRRNLMVDLLVKP
15		190 200 210 220 230 240
	orf112ng	250 260 270 280 290 300
	orf112-1	DQMSVGELTTYIRHLQNNSONTQIYIAIAWRKLVYPVAAWVMAVAFPTPQTRHNGMG
20		250 260 270 280 290 300
	orf112ng	310 320 330 340 350
	orf112-1	LKLFGGICLGLLFLHLAGRLFGFTSOLYGTTPFFLAGALPTIAFALLAVLIRKQEKRX
25		310 320

30 This analysis suggests that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 105

Table III lists several *Neisseria* strains which were used to assess the conservation of the sequence of ORF 4 among different strains.

35 **TABLE III – List of *Neisseria* Strains Used for Gene Variability Study of ORF 4**

ORF4 gene variability: List of used *Neisseria* strains

Identification number	Strains	Source / reference
	Group B	
zv01_4	NG6/88	R. Moxon / Seiler <i>et al.</i> , 1996
zv02_4	BZ198	R. Moxon / Seiler <i>et al.</i> , 1996
zv03_4ass	NG3/88	R. Moxon / Seiler <i>et al.</i> , 1996
zv04_4	297-0	R. Moxon / Seiler <i>et al.</i> , 1996
zv05_4	1000	R. Moxon / Seiler <i>et al.</i> , 1996
zv06_4	BZ147	R. Moxon / Seiler <i>et al.</i> , 1996
zv07_4	BZ169	R. Moxon / Seiler <i>et al.</i> , 1996
zv08_4	528	R. Moxon / Seiler <i>et al.</i> , 1996
zv09_4	NGP165	R. Moxon / Seiler <i>et al.</i> , 1996
zv10_4	BZ133	R. Moxon / Seiler <i>et al.</i> , 1996

zv11_4	NGE31	R. Moxon / Seiler <i>et al.</i> , 1996
zv12_4ass	NGF26	R. Moxon / Seiler <i>et al.</i> , 1996
zv13_4	NGE28	R. Moxon / Seiler <i>et al.</i> , 1996
zv15_4	SWZ107	R. Moxon / Seiler <i>et al.</i> , 1996
zv16_4	NGH15	R. Moxon / Seiler <i>et al.</i> , 1996
zv17_4	NGH36	R. Moxon / Seiler <i>et al.</i> , 1996
zv18_4	BZ232	R. Moxon / Seiler <i>et al.</i> , 1996
zv19_4	BZ83	R. Moxon / Seiler <i>et al.</i> , 1996
zv20_4	44/76	R. Moxon / Seiler <i>et al.</i> , 1996
zv21_4	MC58	R. Moxon
zv96_4	2996	Our collection
Group A		
zv22_4	205900	R. Moxon
z2491_4	Z2491	R. Moxon / Maiden <i>et al.</i> , 1998
Group C		
zv24_4	90/18311	R. Moxon
zv25_4	93/4286	R. Moxon
Others		
zv26_4ass	A22 (group W)	R. Moxon / Maiden <i>et al.</i> , 1998
zv27_4	E26 (group X)	R. Moxon / Maiden <i>et al.</i> , 1998
zv28_4	860800 (group Y)	R. Moxon / Maiden <i>et al.</i> , 1998
zv29_4	E32 (group Z)	R. Moxon / Maiden <i>et al.</i> , 1998
Gonococcus		
zv32_4	Ng F62	R. Moxon / Maiden <i>et al.</i> , 1998
zv33_4	Ng SN4	R. Moxon
fa1090_4	FA1090	R. Moxon
References:		
Seiler A. <i>et al.</i> , Mol. Microbiol., 1996, 19(4):841-856.		
Maiden <i>et al.</i> , Proc. Natl. Acad. Sci. USA, 1998, 95:3140-3145.		

The amino acid sequences for each listed strain are as follows:

>FA1090_4 <SEQ ID 893>
 MWTFFKTLGAALALILAAGGQKDSAPAAASRAAPSDNAGAAKKEIVFGTTVGDFGDMVK
 EQIQALEKKGYTVKLVEFTDYVRPNLALAEGLDINVFQHKFYLLDDKKEHNDITEAF
 QVPTAPLGLYPGKLKSLVEEYKDGSTVSAPNDPSNFARALVMLNELGWIKLKDGINPLTAS

KADIAENLNKNIKIVELEAAQLPRSRADVDFAVVGNYAIISSGMKLTEALFQEPSFAYVNW
SAVKTADKDSOWLKDVTAYNSDAFKAYAHKREFEGYKYPAAWNEGAAK*

5 >Z2491.4 <SEQ ID 894>
MKTFITFLSAAALAILAAGCGGQKSDAPAAASAAADNGAAKKEIVFGTTVGDFGDMVKE
QIQPLEKKGYTVKLVEFTYDVPNRNLALAEGLDINVFQHKPYLDDFKKEHNLDTVEVQ
VPTAPLGLYPGKLKSLSEEVKDGSTVSAPNDPSNFARVLVMLDELGLWIKLKGINPLTASK
ADTAENLKNIKIWLAEAYPLPRSRADDFVAVVNGNYAISGSMKLTALFQEPSFAVNVNS
AVKTADEKDKCMLKDVTEAASQDAFKAYAHKFEQGYKSPAAWNGAAK*

10 AVATADKDSQWLEWVLEATINSLAKATANKREKESQVPAWNGEAK
 >ZV01_4 <SEQ ID 895>
 MKTFFKTLASAAALALILACGGQKDSAPAASASAAADNGAAKEIVFGTTVGDFGDMVKE
 VQLELEKKGYTKVLFEEVTVRNPLALAEGLDINVFQHPKVLIDDFKKEHNLDTIEFVQ
 15 VPTAPLGLYPGKLGKLETTVDKGSSTVSPANDSNFARVLVLMDELGWIKLGDINPLTASQ
 ADAENLNKRIIVLEEAQLPSRAUDVFAVNGNGYIALSSGGMKLTALFEAPSPAYNVNS
 AVKTDKNSOWLKDVTAYNSDAFKAIAHRRFGEYKSPAAWNEGAAK*

20 >zv02_4 (SEQ ID 896):
MKTFFETKLSSAALAILAACGGQKSDAPASASAAADNGAEKKEIVFGTTVGDFGDMVKE
HIQPELEKKGYTVKLVEFTDYVRPNLALAEGLDINVQHPKPYLDDFKKEHNLDTVEFG
VTPPLGLGYFGKLKSLSEEVKDGSTVSAPNDPSNFARVLVLMDELGWIKLKGINPLTASK
ADAETLENKIKIWLAEAAQLPSRADYDAFVAVNGYKSSGMRLTEALFQEPSFAYVNS
AVKTDKDSOSMLKDVTEAYNSDAFVAHKKRFGGYKSPPAENGAA*

25 >VZ03_4ASS <SEQ ID 897>
MKTFFTKTSAARALLIALAACGGQKDSAPAASASAADNAGSKEKKEIVFGTTVGGDFGDMVKE
HIQPELEKKQYTVKLVFTYVYRPNLALAEGLDINVPQHPRYLDGFKKEHNLDITEVQF
VTPAPLGLYFGKLSLEEVKDGSTVSAPNDPSNFARVLVMLDELGWIKLKDGINPLTASK
30 ADAENLNKNIKILEAAQLPRSDADYFVAVNGNYAISGKMLTEALFQEPSFAYVNW
ADTAKCSOWLKDVTETAPKADCAIYAKHRRVEGYSPPAANNMGA*

35 >ZV04_4 <SEQ ID 898>
MKTFFELSLAALALILACGGQKSDSAPAASASAADNGAEKKEIVFGTTVGDGDFGMVKE
HIQPELEKKGYTVKLVFTDVPYRNLALAEGLDINVQHPKPYLDDFKKEHNLDITVFPQ
VTPATGLGYFGFKLSLEEVDKSTVSAPNDPSNFARVLVMLDELGWIKLKDGINFLTASK
ADAENLNKNIHILEAAGLPDSADFVAVNGNYAISPAAGLKALFQEPSPYAVNMS
AVKATADKDSOWLKDVTETAYNSDRAKYAHKRRFEGYKSPAAWNGKA*

40 >ZV05_4 <SEQ ID 899>
MKTFTFLSAAALALAILAACGGQKSDSAPAAASAADNGAEKKEIVFGTITVGDFDGMVKE
HIQPELEKKGYTVKLVEFTDYVRPNLALAEGLDINVFQHPYLDLDFKKEHNLDTVEVQ
VPTAPLGLYPGKLKSLBEVKDGSVTSAPNDFSNFARVLVMLDELGWIKLKGINFLTASK
ADIAENLNKIKIVLEAAQLPRSDVDFVAVNGEYSISPMGLTALFQFSPFAYVNS
AVKTDKDSQWLKDVTEA VNSDAPKAYHRKFGYKSAAGKNNGEAA*

50 >ZV06_4 <SEQ ID 900>
MKTFFKTSAAALALILACGGCKDSAPAAASAAADNGAEKKEIVFGTTVGDGDFGMVKE
QIQAELEKFGYTKLVVEFTDYVRPNLALAEGLDINVFQHPYLDGDFKKNHLDITVFPQ
VPTAPLGLYPGKLSLEEVKDGSTVSAPNDPSNFRVILVIMLDELGWIKLKGINPLTASK
ADIAENLNKIKIVLEAAQLPRSRADGDFAVVNGNYISGGMKLTEALFQFSPFAYVNW
AVTKAHKDSQWLKDVDTVEAYNSDAFKAHYHKRFEGYKSPAAENNEGA*

55 >ZV07.4 <SEQ ID 901>
MKTFFKTSAAALAILAACGGQKSDSAPAAASAAADNGAAKKEIVFGTTVGDFDGMVKE
QIQAELEKFGYTKLVFFETDVTYRPNLALAEGLDINVFQHKPYLDFFKEHNLDITVFP
VPTAPLGLYPGKLSLEEVKDGSTVSAPNDFSNFARVILVMLDELGWIKLKGINPLTAK
ADIAENLNKIKIVLEAAQLPRASRDVFVAVNGNYAISGGMKLTEALFQEPSFAYVNW
AVTKADKDSQWLKDVETYSNDAFKAYAKRFEGYKSPAANKNEGAAK*

60 >ZV08_4 <SEQ ID 902>
 MKTFEFTKLSAALALILVAFTCGGQKSDSAPAASASAAAANGAAKEKKEIVFGTTVGDGDFGMVKE
 HIQPELEKKSQTVELVFQDYVVRPNLALAEGLDINVFQHPYPLDQDFKEHNLDITEVQ
 VPTAPLGLYPGKLSLEEVKDGSTVSAPNDPSNFARVLVMLDELGWIKLKDGINPLTASK
 ADIAENLNKIKIVLEEAQLPSRSDVDFVAVNGYKSSPMGKLTALFQEPSFAYVNW
 65 AQLTADKDSQWLKDVTETAYSNDAPKAHYHREFGYKSPAAANGAA*

70 >ZV09_4 <SEQ ID 902>
MKTFTLSAALAALILAACGGQKSDSAPAASRSAAADNGAEKKEIVFGTTVGDGDFGMVKE
HIQPELEKFTGYTVKLFVFTDYVRNPLALAEGLDINVPQHFKPYLDDFKKEHNLDITVEPQ
VPTAPLGLYPGKLSLEEVDKSTVSAPNDPSNFARVVLMLDELGWIKLKDGINPLTASK
ADIAENLNKIKIVLEAALGPLRSRADYFVAVNGNYASPMNGKLTALFQEPSFAYDNWS
AVKTAQKQISWLKDVTEAYNSDAKCAVYHKRGYKSPAAWNGEAA*

75 >zv10_4<SEQ ID 903>
MKTFFTSLSAALALILAACGGQKSDSAPAASASAAADNGAAKKEIVFGTTVGDGDMVKE
HIQPELEKKGTVKLVFETVYRPNLALAEGLDINVQHPKPYLDQFKEHNLDITEVQF
VPTAPLGLYPGKLSLEEVKDGSTVSAPNDPSNFARVLVMLDELGWIKLDGINPLTASK
DAIAENLNKIKIVLEAAQLPRSRADVFVAVNGNYKSSGMRLTEALFQEPSFAVWNVS
AVKTADYKDSWLKDVTEAYNSKDAFYAHRKFEYGKSPAAWNEGAA*

>ZV11_4 <SEQ ID 904>
MKTFFKTLAALALILAAACGGQKDSAPAASASAAADNGAAKKEIVFGTTVGDFGDMVKE
QIQAELEKKGTYTKLVFTDYVRPNLALAEGLDINVFQHKPYLDDFKKEHNLDTIEVFQ
VPTAPLGLYPGKLKSLSEEVKDGSTVSA PNDPSNFARVLVLMDELGWIKLKDGINPLTASK
ADIAENLNKIKIVELEAAQLPRSRADVFAVNGNYAISSGMKLTEALFQEPSFAYVNW
AVKTDKDSQWLKDVTEAYNSDAFKAYAHKRFEGYKSPAAWNEGAAK*

>ZV12_4ASS <SEQ ID 905>
MKTFFKTLAALALILAAACGGQKDRAPAAASAAASENGAAKKEILFGTTVGDGLDMVKE
QIQAELEKKGTYTKLVFTDYVRPNLALAEGLDINVFQHKPYLDDFKKEHNLDTIEVFQ
VPTAPLGLYPGKLKSLSEEVKDGSTVSA PNDPSNFARVLVLMDELGWIKLKDGINPLTASK
ADIAENLNKIKIVELEAAQLPRSRADVFAVNGNYAISSGMKLTEALFQEPSFAYVNW
AVKTDKDSQWLKDVTEAYNSDAFKAYAHKRFEGYKSPAAWNEGAAK*

>ZV13_4 <SEQ ID 906>
MKTFFKTLAALALILAAACGGQKDSAPAASASAAADNGAAKKEIVFGTTVGDFGDMVKE
QIQAELEKKGTYTKLVFTDYVRPNLALAEGLDINVFQHKPYLDDFKKEHNLDTIEVFQ
VPTAPLGLYPGKLKSLSEEVKDGSTVSA PNDPSNFARVLVLMDELGWIKLKDGINPLTASK
ADIAENLNKIKIVELEAAQLPRSRADVFAVNGNYAISSGMKLTEALFQEPSFAYVNW
AVKTDKDSQWLKDVTEAYNSDAFKAYAHKRFEGYKSPAAWNEGAAK*

>ZV15_4 <SEQ ID 907>
MKTFFKTLAALALILAAACGGQKDSAPAASASAAADNGAAKKEIVFGTTVGDFGDMVKE
HIQPELEKKGTYTKLVFTDYVRPNLALAEGLDINVFQHKPYLDDFKKEHNLDTIEVFQ
VPTAPLGLYPGKLKSLSEEVKDGSTVSA PNDPSNFARVLVLMDELGWIKLKDGINPLTASK
ADIAENLNKIKIVELEAAQLPRSRADVFAVNGNYAISSGMKLTEALFQEPSFAYVNW
AVKTDKDSQWLKDVTEAYNSDAFKAYAHKRFEGYKSPAAWNEGAAK*

>ZV16_4 <SEQ ID 908>
MKTFFKTLAALALILAAACGGQKDSAPAASASAAADNGAAKKEIVFGTTVGDFGDMVKE
HIQPELEKKGTYTKLVFTDYVRPNLALAEGLDINVFQHKPYLDDFKKEHNLDTIEVFQ
VPTAPLGLYPGKLKSLSEEVKDGSTVSA PNDPSNFARVLVLMDELGWIKLKDGINPLTASK
ADIAENLNKIKIVELEAAQLPRSRADVFAVNGNYAISSGMKLTEALFQEPSFAYVNW
AVKTDKDSQWLKDVTEAYNSDAFKAYAHKRFEGYKSPAAWNEGAAK*

>ZV17_4 <SEQ ID 909>
MKTFFKTLAALALILAAACGGQKDSAPAASASAAADNGAAKKEIVFGTTVGDFGDMVKE
QIQAELEKKGTYTKLVFTDYVRPNLALAEGLDINVFQHKPYLDDFKKEHNLDTIEVFQ
VPTAPLGLYPGKLKSLSEEVKDGSTVSA PNDPSNFARVLVLMDELGWIKLKDGINPLTASK
ADIAENLNKIKIVELEAAQLPRSRADVFAVNGNYAISSGMKLTEALFQEPSFAYVNW
AVKTDKDSQWLKDVTEAYNSDAFKAYAHKRFEGYKSPAAWNEGAAK*

>ZV18_4 <SEQ ID 910>
MKTFFKTLAALALILAAACGGQKDSAPAASASAAADNGAAKKEIVFGTTVGDFGDMVKE
HIQPELEKKGTYTKLVFTDYVRPNLALAEGLDINVFQHKPYLDDFKKEHNLDTIEVFQ
VPTAPLGLYPGKLKSLSEEVKDGSTVSA PNDPSNFARVLVLMDELGWIKLKDGINPLTASK
ADIAENLNKIKIVELEAAQLPRSRADVFAVNGNYAISSGMKLTEALFQEPSFAYVNW
AVKTDKDSQWLKDVTEAYNSDAFKAYAHKRFEGYKSPAAWNEGAAK*

>ZV19_4 <SEQ ID 911>
MKTFFKTLAALALILAAACGGQKDSAPAASASAAADNGAAKKEIVFGTTVGDFGDMVKE
QIQAELEKKGTYTKLVFTDYVRPNLALAEGLDINVFQHKPYLDDFKKEHNLDTIEVFQ
VPTAPLGLYPGKLKSLSEEVKDGSTVSA PNDPSNFARVLVLMDELGWIKLKDGINPLTASK
ADIAENLNKIKIVELEAAQLPRSRADVFAVNGNYAISSGMKLTEALFQEPSFAYVNW
AVKTDKDSQWLKDVTEAYNSDAFKAYAHKRFEGYKSPAAWNEGAAK*

>ZV20_4 <SEQ ID 912>
MKTFFKTLAALALILAAACGGQKDSAPAASASAAADNGAAKKEIVFGTTVGDFGDMVKE
QIQAELEKKGTYTKLVFTDYVRPNLALAEGLDINVFQHKPYLDDFKKEHNLDTIEVFQ
VPTAPLGLYPGKLKSLSEEVKDGSTVSA PNDPSNFARVLVLMDELGWIKLKDGINPLTASK
ADIAENLNKIKIVELEAAQLPRSRADVFAVNGNYAISSGMKLTEALFQEPSFAYVNW
AVKTDKDSQWLKDVTEAYNSDAFKAYAHKRFEGYKSPAAWNEGAAK*

>ZV21_4 <SEQ ID 913>
MKTFFKTLAALALILAAACGGQKDSAPAASASAAADNGAAKKEIVFGTTVGDFGDMVKE
QIQAELEKKGTYTKLVFTDYVRPNLALAEGLDINVFQHKPYLDDFKKEHNLDTIEVFQ
VPTAPLGLYPGKLKSLSEEVKDGSTVSA PNDPSNFARVLVLMDELGWIKLKDGINPLTASK
ADIAENLNKIKIVELEAAQLPRSRADVFAVNGNYAISSGMKLTEALFQEPSFAYVNW
AVKTDKDSQWLKDVTEAYNSDAFKAYAHKRFEGYKSPAAWNEGAAK*

>ZV22_4 <SEQ ID 914>
MKTFFKTLAALALILAAACGGQKDSAPAASASAAADNGAAKKEIVFGTTVGDFGDMVKE
QIQAELEKKGTYTKLVFTDYVRPNLALAEGLDINVFQHKPYLDDFKKEHNLDTIEVFQ
VPTAPLGLYPGKLKSLSEEVKDGSTVSA PNDPSNFARVLVLMDELGWIKLKDGINPLTASK
ADIAENLNKIKIVELEAAQLPRSRADVFAVNGNYAISSGMKLTEALFQEPSFAYVNW
AVKTDKDSQWLKDVTEAYNSDAFKAYAHKRFEGYKSPAAWNEGAAK*

>ZV24_4ASS <SEQ ID 915>

MKTFFKTLISAALALILAACGGQKDSAPAASASAAADNGAEKKEIVFGTTVGDFGDMVKE
 HIQPELEKKGYYTVKLVFTDYVRPNLALAEGLDINVFQHKPYLDDFKKEHNLDTIEVFQ
 VPTAPLGLYPGKLKSLSEEVKDGSTVSAFNDPSNFARVLVMDLGLWIKLKGGINPLTASK
 ADIAENLNKIKIVELEAAQLPSRADVDFAVNGNYAIISSGMKLTALFQEPSFAYVNW
 AVKTADKDSQWLKDVTEAYNSDAFKAYAHKREFEGYKSPAANWEGAAK*

>ZV25_4 <SEQ ID 916>

MKTFFKTLISAALALILAACGGQKDSAPAASASAAADNGAEKKEIVFGTTVGDFGDMVKE
 QIQPELEKKGYYTVKLVFTDYVRPNLALAEGLDINVFQHKPYLDDFKKEHNLDTIEVFQ
 VPTAPLGLYPGKLKSLSEEVKDGSTVSAFNDPSNFARVLVMDLGLWIKLKGGINPLTASK
 ADIAENLNKIKIVELEAAQLPSRADVDFAVNGNYAIISSGMKLTALFQEPSFAYVNW
 AVKTADKDSQWLKDVTEAYNSDAFKAYAHKREFEGYKSPAANWEGAAK*

>ZV26_4 <SEQ ID 917>

MKTFFKTLISAALALILAACGGQKDSAPAASASAAADNGAEKKEIVFGTTVGDFGDMVKE
 HIQPELEKKGYYTVKLVFTDYVRPNLALAEGLDINVFQHKPYLDDFKKEHNLDTIEVFQ
 VPTAPLGLYPGKLKSLSEEVKDGSTVSAFNDPSNFARVLVMDLGLWIKLKGGINPLTASK
 ADIAENLNKIKIVELEAAQLPSRADVDFAVNGNYAIISSGMKLTALFQEPSFAYVNW
 AVKTADKDSQWLKDVTEAYNSDAFKAYAHKREFEGYKSPAANWEGAAK*

>ZV27_4 <SEQ ID 918>

MKTFFKTLISAALALILAACGGQKDSAPAASASAAADNGAEKKEIVFGTTVGDFGDMVKE
 QIQPELEKKGYYTVKLVFTDYVRPNLALAEGLDINVFQHKPYLDDFKKEHNLDTIEVFQ
 VPTAPLGLYPGKLKSLSEEVKDGSTVSAFNDPSNFARVLVMDLGLWIKLKGGINPLTASK
 ADIAENLNKIKIVELEAAQLPSRADVDFAVNGNYAIISSGMKLTALFQEPSFAYVNW
 AVKTADKDSQWLKDVTEAYNSDAFKAYAHKREFEGYKSPAANWEGAAK*

>ZV28_4 <SEQ ID 919>

MKTFFKTLISAALALILAACGGQKDSAPAASASAAADNGAEKKEIVFGTTVGDFGDMVKE
 HIQPELEKKGYYTVKLVFTDYVRPNLALAEGLDINVFQHKPYLDDFKKEHNLDTIEVFQ
 VPTAPLGLYPGKLKSLSEEVKDGSTVSAFNDPSNFARVLVMDLGLWIKLKGGINPLTASK
 ADIAENLNKIKIVELEAAQLPSRADVDFAVNGNYAIISSGMKLTALFQEPSFAYVNW
 AVKTADKDSQWLKDVTEAYNSDAFKAYAHKREFEGYKSPAANWEGAAK*

>ZV29_4 <SEQ ID 920>

MKTFFKTLISAALALILAACGGQKDSAPAASASAAADNGAEKKEIVFGTTVGDFGDMVKE
 QIQPELEKKGYYTVKLVFTDYVRPNLALAEGLDINVFQHKPYLDDFKKEHNLDTIEVFQ
 VPTAPLGLYPGKLKSLSEEVKDGSTVSAFNDPSNFARVLVMDLGLWIKLKGGINPLTASK
 ADIAENLNKIKIVELEAAQLPSRADVDFAVNGNYAIISSGMKLTALFQEPSFAYVNW
 AVKTADKDSQWLKDVTEAYNSDAFKAYAHKREFEGYKSPAANWEGAAK*

>ZV32_4 <SEQ ID 921>

MKTFFKTLISAALALILAACGGQKDSAPAASASAAADNGAEKKEIVFGTTVGDFGDMVKE
 EQIQAELEKKGYYTVKLVFTDYVRPNLALAEGLDINVFQHKPYLDDFKKEHNLDTIEVFQ
 QVPTAPLGLYPGKLKSLSEEVKDGSTVSAFNDPSNFARVLVMDLGLWIKLKGGINPLTASK
 KADIAENLNKIKIVELEAAQLPSRADVDFAVNGNYAIISSGMKLTALFQEPSFAYVNW
 SAVKTADKDSQWLKDVTEAYNSDAFKAYAHKREFEGYKSPAANWEGAAK*

>ZV33_4 <SEQ ID 922>

MKTFFKTLISAALALILAACGGQKDSAPAASASAAADNGAEKKEIVFGTTVGDFGDMVKE
 EQIQAELEKKGYYTVKLVFTDYVRPNLALAEGLDINVFQHKPYLDDFKKEHNLDTIEVFQ
 QVPTAPLGLYPGKLKSLSEEVKDGSTVSAFNDPSNFARVLVMDLGLWIKLKGGINPLTASK
 KADIAENLNKIKIVELEAAQLPSRADVDFAVNGNYAIISSGMKLTALFQEPSFAYVNW
 SAVKTADKDSQWLKDVTEAYNSDAFKAYAHKREFEGYKSPAANWEGAAK*

>ZV96_4 <SEQ ID 923>

MKTFFKTLISAALALILAACGGQKDSAPAASASAAADNGAEKKEIVFGTTVGDFGDMVKE
 QIQAELEKKGYYTVKLVFTDYVRPNLALAEGLDINVFQHKPYLDDFKKEHNLDTIEVFQ
 VPTAPLGLYPGKLKSLSEEVKDGSTVSAFNDPSNFARVLVMDLGLWIKLKGGINPLTASK
 ADIAENLNKIKIVELEAAQLPSRADVDFAVNGNYAIISSGMKLTALFQEPSFAYVNW
 AVKTADKDSQWLKDVTEAYNSDAFKAYAHKREFEGYKSPAANWEGAAK*

Figure 8 shows the results of aligning the sequences of each of these strains. Dark shading indicates regions of homology, and gray shading indicates the conservation of amino acids with similar characteristics. As is readily discernible, there is significant conservation among the various strains of ORF 4, further confirming its utility as an antigen for both vaccines and diagnostics.

It will be appreciated that the invention has been described by means of example only, and that modifications may be made whilst remaining within the spirit and scope of the invention.

60210-8120360

CLAIMS

1. A protein comprising an amino acid sequence selected from the group consisting of SEQ IDs 2, 4, 6, and 8.

2. A nucleic acid molecule which encodes a protein according to claim 1.

3. A nucleic acid molecule according to claim 2, comprising a nucleotide sequence selected from the group consisting of SEQ IDs 1, 3, 5, and 7.

4. A protein comprising an amino acid sequence selected from the group consisting of SEQ IDs 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252, 254, 256, 258, 260, 262, 264, 266, 268, 270, 272, 274, 276, 278, 280, 282, 284, 286, 288, 290, 292, 294, 296, 298, 300, 302, 304, 306, 308, 310, 312, 314, 316, 318, 320, 322, 324, 326, 328, 330, 332, 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370, 372, 374, 376, 378, 380, 382, 384, 386, 388, 390, 392, 394, 396, 398, 400, 402, 404, 406, 408, 410, 412, 414, 416, 418, 420, 422, 424, 426, 428, 430, 432, 434, 436, 438, 440, 442, 444, 446, 448, 450, 452, 454, 456, 458, 460, 462, 464, 466, 468, 470, 472, 474, 476, 478, 480, 482, 484, 486, 488, 490, 492, 494, 496, 498, 500, 502, 504, 506, 508, 510, 512, 514, 516, 518, 520, 522, 524, 526, 528, 530, 532, 534, 536, 538, 540, 542, 544, 546, 548, 550, 552, 554, 556, 558, 560, 562, 564, 566, 568, 570, 572, 574, 576, 578, 580, 582, 584, 586, 588, 590, 592, 594, 596, 598, 600, 602, 604, 606, 608, 610, 612, 614, 616, 618, 620, 622, 624, 626, 628, 630, 632, 634, 636, 638, 640, 642, 644, 646, 648, 650, 652, 654, 656, 658, 660, 662, 664, 666, 668, 670, 672, 674, 676, 678, 680, 682, 684, 686, 688, 690, 692, 694, 696, 698, 700, 702, 704, 706, 708, 710, 712, 714, 716, 718, 720, 722, 724, 726, 728, 730, 732, 734, 736, 738, 740, 742, 744, 746, 748, 750, 752, 754, 756, 758, 760, 762, 764, 766, 768, 770, 772, 774, 776, 778, 780, 782, 784, 786, 788, 790, 792, 794, 796, 798, 800, 802, 804, 806, 808, 810, 812, 814, 816, 818, 820, 822, 824, 826, 828, 830, 832, 834, 836, 838, 840, 842, 844, 846, 848, 850, 852, 854, 856, 858, 860, 862, 864, 866, 868, 870, 872, 874, 876, 878, 880, 882, 884, 886, 888, 890, & 892..

5. A protein having 50% or greater sequence identity to a protein according to claim 4.

6. A protein comprising a fragment of an amino acid sequence selected from the group consisting of SEQ IDs 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252, 254, 256, 258, 260, 262, 264, 266, 268, 270, 272, 274, 276, 278, 280, 282, 284, 286, 288, 290, 292, 294, 296, 298, 300, 302, 304, 306, 308, 310, 312, 314, 316, 318, 320, 322, 324, 326, 328, 330, 332, 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370, 372, 374, 376, 378, 380, 382, 384, 386, 388, 390, 392, 394, 396, 398, 400, 402, 404, 406, 408, 410, 412, 414, 416, 418, 420, 422, 424, 426, 428, 430, 432, 434, 436, 438, 440, 442, 444, 446, 448, 450, 452, 454, 456, 458, 460, 462, 464, 466, 468, 470, 472, 474, 476, 478, 480, 482, 484, 486, 488, 490, 492, 494, 496, 498, 500, 502, 504, 506, 508, 510, 512, 514, 516, 518, 520, 522, 524, 526, 528, 530, 532, 534, 536, 538, 540, 542, 544, 546, 548, 550, 552, 554, 556, 558, 560, 562, 564, 566, 568, 570, 572, 574, 576, 578, 580, 582, 584, 586, 588, 590, 592, 594, 596, 598, 600, 602, 604, 606, 608, 610, 612, 614, 616, 618, 620, 622, 624, 626, 628, 630, 632, 634, 636, 638, 640, 642, 644, 646, 648, 650, 652, 654, 656, 658, 660, 662, 664, 666, 668, 670, 672, 674, 676, 678, 680, 682, 684, 686, 688, 690, 692, 694, 696, 698, 700, 702, 704, 706, 708, 710, 712, 714, 716, 718, 720, 722, 724, 726, 728, 730, 732, 734, 736, 738, 740, 742, 744, 746, 748, 750, 752, 754, 756, 758, 760, 762, 764, 766, 768, 770, 772, 774, 776, 778, 780, 782, 784, 786, 788, 790, 792, 794, 796, 798, 800, 802, 804, 806, 808, 810, 812, 814, 816, 818, 820, 822, 824, 826, 828, 830, 832, 834, 836, 838, 840, 842, 844, 846, 848, 850, 852, 854, 856, 858, 860, 862, 864, 866, 868, 870, 872, 874, 876, 878, 880, 882, 884, 886, 888, 890, & 892..

7. An antibody which binds to a protein according to any one of claims 4 to 6.

8. A nucleic acid molecule which encodes a protein according to any one of claims 4 to 6.

9. A nucleic acid molecule according to claim 8, comprising a nucleotide sequence selected from the group consisting of SEQ IDs 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249,

251, 253, 255, 257, 259, 261, 263, 265, 267, 269, 271, 273, 275, 277, 279, 281, 283, 285, 287, 289,
291, 293, 295, 297, 299, 301, 303, 305, 307, 309, 311, 313, 315, 317, 319, 321, 323, 325, 327, 329,
331, 333, 335, 337, 339, 341, 343, 345, 347, 349, 351, 353, 355, 357, 359, 361, 363, 365, 367, 369,
371, 373, 375, 377, 379, 381, 383, 385, 387, 389, 391, 393, 395, 397, 399, 401, 403, 405, 407, 409,
5 411, 413, 415, 417, 419, 421, 423, 425, 427, 429, 431, 433, 435, 437, 439, 441, 443, 445, 447, 449,
451, 453, 455, 457, 459, 461, 463, 465, 467, 469, 471, 473, 475, 477, 479, 481, 483, 485, 487, 489,
491, 493, 495, 497, 499, 501, 503, 505, 507, 509, 511, 513, 515, 517, 519, 521, 523, 525, 527, 529,
531, 533, 535, 537, 539, 541, 543, 545, 547, 549, 551, 553, 555, 557, 559, 561, 563, 565, 567, 569,
571, 573, 575, 577, 579, 581, 583, 585, 587, 589, 591, 593, 595, 597, 599, 601, 603, 605, 607, 609,
10 611, 613, 615, 617, 619, 621, 623, 625, 627, 629, 631, 633, 635, 637, 639, 641, 643, 645, 647, 649,
651, 653, 655, 657, 659, 661, 663, 665, 667, 669, 671, 673, 675, 677, 679, 681, 683, 685, 687, 689,
691, 693, 695, 697, 699, 701, 703, 705, 707, 709, 711, 713, 715, 717, 719, 721, 723, 725, 727, 729,
731, 733, 735, 737, 739, 741, 743, 745, 747, 749, 751, 753, 755, 757, 759, 761, 763, 765, 767, 769,
771, 773, 775, 777, 779, 781, 783, 785, 787, 789, 791, 793, 795, 797, 799, 801, 803, 805, 807, 809,
15 811, 813, 815, 817, 819, 821, 823, 825, 827, 829, 831, 833, 835, 837, 839, 841, 843, 845, 847, 849,
851, 853, 855, 857, 859, 861, 863, 865, 867, 869, 871, 873, 875, 877, 879, 881, 883, 885, 887, 889,
& 891..

10. A nucleic acid molecule comprising a fragment of a nucleotide sequence selected from the
group consisting of SEQ IDs 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39,
20 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91,
93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133,
135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173,
175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213,
215, 217, 219, 221, 223, 225, 227, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253,
25 255, 257, 259, 261, 263, 265, 267, 269, 271, 273, 275, 277, 279, 281, 283, 285, 287, 289, 291, 293,
295, 297, 299, 301, 303, 305, 307, 309, 311, 313, 315, 317, 319, 321, 323, 325, 327, 329, 331, 333,
335, 337, 339, 341, 343, 345, 347, 349, 351, 353, 355, 357, 359, 361, 363, 365, 367, 369, 371, 373,
375, 377, 379, 381, 383, 385, 387, 389, 391, 393, 395, 397, 399, 401, 403, 405, 407, 409, 411, 413,
415, 417, 419, 421, 423, 425, 427, 429, 431, 433, 435, 437, 439, 441, 443, 445, 447, 449, 451, 453,
30 455, 457, 459, 461, 463, 465, 467, 469, 471, 473, 475, 477, 479, 481, 483, 485, 487, 489, 491, 493,
495, 497, 499, 501, 503, 505, 507, 509, 511, 513, 515, 517, 519, 521, 523, 525, 527, 529, 531, 533,
535, 537, 539, 541, 543, 545, 547, 549, 551, 553, 555, 557, 559, 561, 563, 565, 567, 569, 571, 573,
575, 577, 579, 581, 583, 585, 587, 589, 591, 593, 595, 597, 599, 601, 603, 605, 607, 609, 611, 613,
615, 617, 619, 621, 623, 625, 627, 629, 631, 633, 635, 637, 639, 641, 643, 645, 647, 649, 651, 653,

655, 657, 659, 661, 663, 665, 667, 669, 671, 673, 675, 677, 679, 681, 683, 685, 687, 689, 691, 693, 695, 697, 699, 701, 703, 705, 707, 709, 711, 713, 715, 717, 719, 721, 723, 725, 727, 729, 731, 733, 735, 737, 739, 741, 743, 745, 747, 749, 751, 753, 755, 757, 759, 761, 763, 765, 767, 769, 771, 773, 775, 777, 779, 781, 783, 785, 787, 789, 791, 793, 795, 797, 799, 801, 803, 805, 807, 809, 811, 813, 815, 817, 819, 821, 823, 825, 827, 829, 831, 833, 835, 837, 839, 841, 843, 845, 847, 849, 851, 853, 855, 857, 859, 861, 863, 865, 867, 869, 871, 873, 875, 877, 879, 881, 883, 885, 887, 889, & 891..

11. A nucleic acid molecule comprising a nucleotide sequence complementary to a nucleic acid molecule according to any one of claims 8 to 10.

10 12. A nucleic acid molecule comprising a nucleotide sequences having 50% or greater sequence identity to a nucleic acid molecule according to any one of claims 8-11.

13. A nucleic acid molecule which can hybridise to a nucleic acid molecule according to any one of claims 8-12 under high stringency conditions.

14. A composition comprising a protein, a nucleic acid molecule, or an antibody according to any preceding claim.

15 15. A composition according to claim 14 being a vaccine composition or a diagnostic composition.

16. A composition according to claim 14 or claim 15 for use as a pharmaceutical.

17. The use of a composition according to claim 14 in the manufacture of a medicament for the treatment or prevention of infection due to Neisserial bacteria.

ABSTRACT

The invention provides proteins from *Neisseria meningitidis* (strains A & B) and from *Neisseria gonorrhoeae*, including amino acid sequences, the corresponding nucleotide sequences, expression data, and serological data. The proteins are useful antigens for vaccines, immunogenic compositions, and/or diagnostics.

5

09302518.043009

FIGURE 1

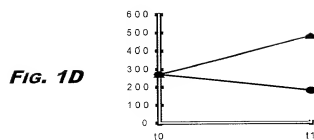
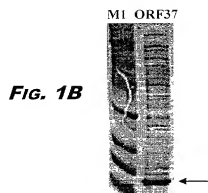
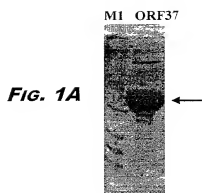


Fig 1E

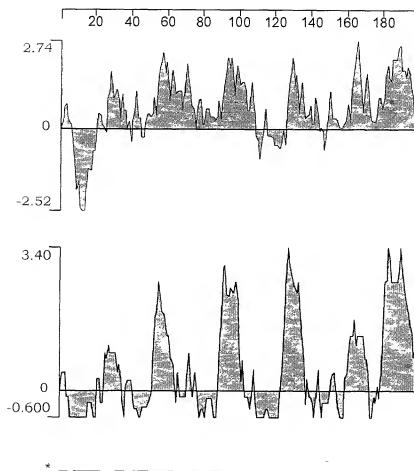


FIGURE 2

FIG. 2A

M1 ORF5



FIG. 2B

TP



FIGURE 3

Fig. 3A

M1 ORF2



Fig. 3B

M1 ORF2

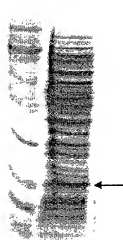


Fig. 3C

TP OMV

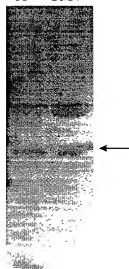
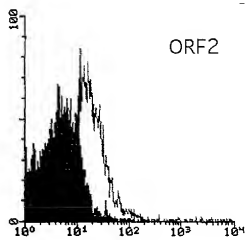


Fig. 3D



09303518-043009

FIGURE 4

Fig. 4A

M1 ORF15

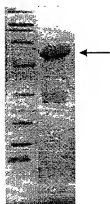


Fig. 4B

M2 ORF15

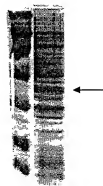


Fig 4C

TP OMV

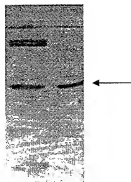


FIGURE 5

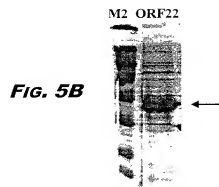
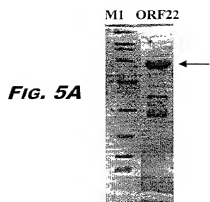
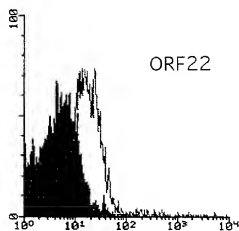


Fig. 5C



00302510.043000

FIGURE 6

Fig. 6A

M1 ORF28

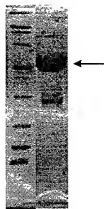


Fig. 6B

M2 ORF28

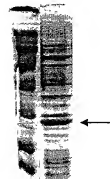


FIGURE 7

Fig. 7A

M1 ORF32

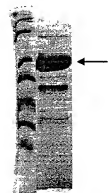


Fig. 7B

M1 ORF32

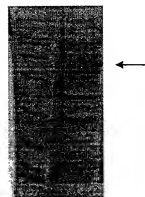


FIGURE 8

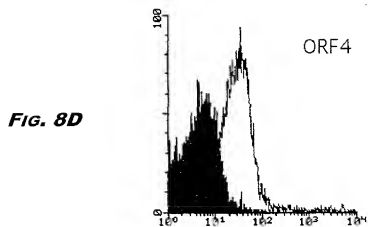
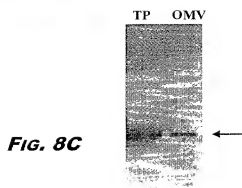
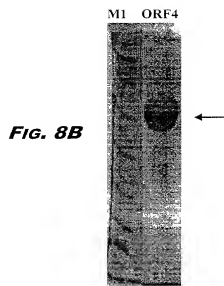
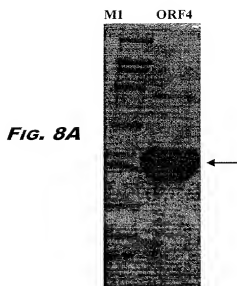


FIG. 8E

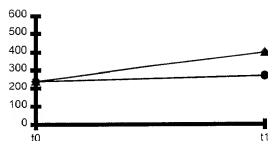
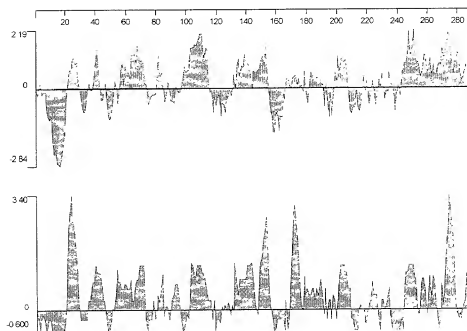
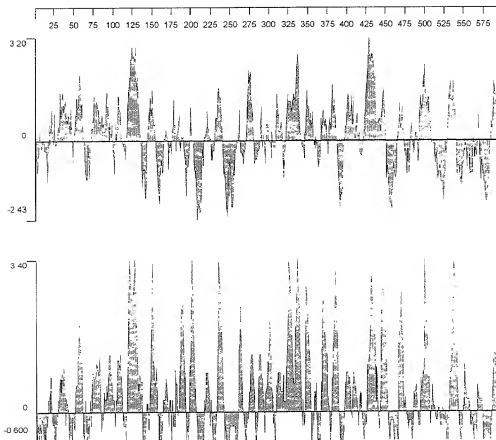


FIG. 8F



00303210-043009

FIGURE 9



600340-8752060

FIGURE 10

Fig. 10A



Fig. 10B

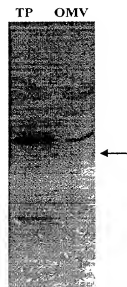


Fig. 10C

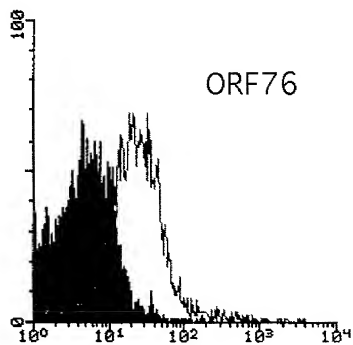


FIGURE 12

FIG. 12A



FIG. 12B

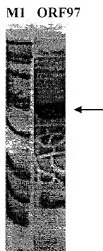


FIG. 12C



FIG. 12D

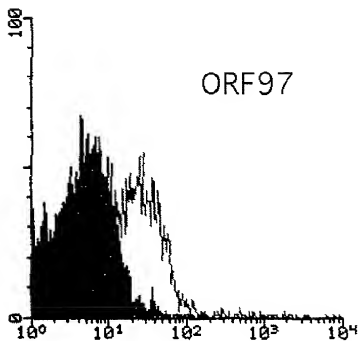


FIGURE 13

Fig. 13A

M1 ORF106



Fig. 13B

M2 ORF106



Fig. 13C

ORF 106

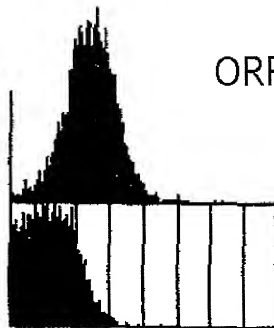


FIGURE 14

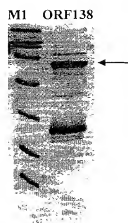


FIG. 14A

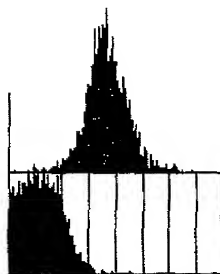


FIG. 14B

FIGURE 15

Fig. 15A

M1 ORF23



Fig. 15B

M2 ORF23



Fig 15C

TP OMV

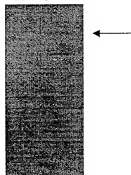


FIGURE 16

FIG. 16A

M2 ORF25

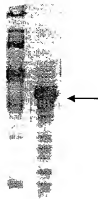


FIG. 16B

M1 ORF25



FIG. 16C

TP OMV



FIG. 16D

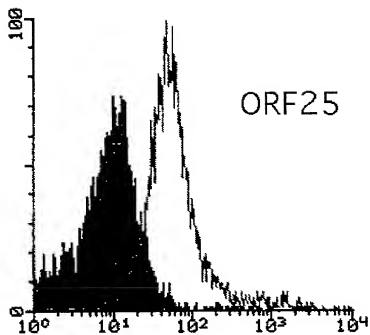
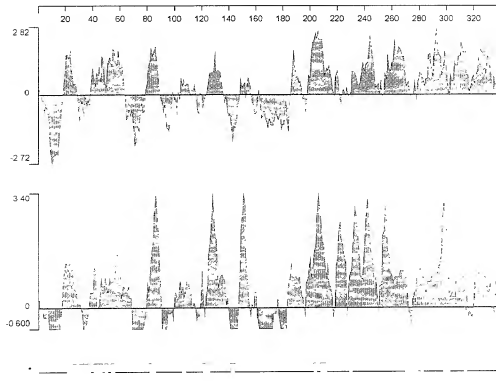


Fig. 16E



00303218-043009

FIGURE 17

FIG. 17A

M1 ORF27

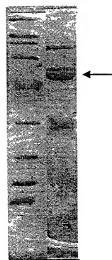


FIG. 17B

M2 ORF27

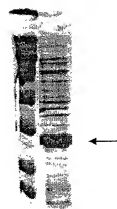


FIGURE 18

FIG. 18A



FIG. 18B

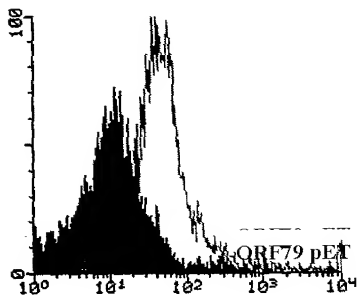


FIGURE 19

Fig. 19A



Fig. 19B

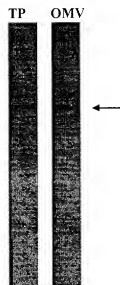


Fig. 19C

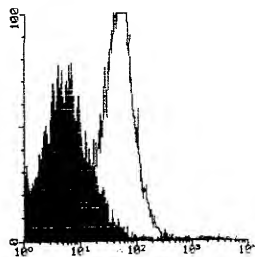
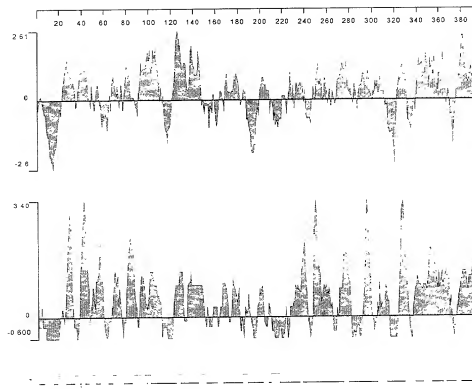


Fig 19D



0000518 043000

FIGURE 20

Fig. 20A

M1 ORF132

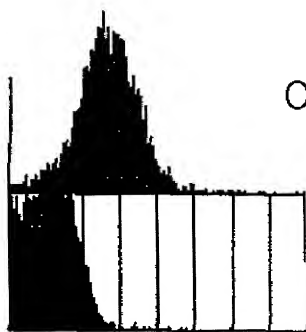


Fig. 20B

M2 ORF132



Fig. 20C



ORF132


```

fa1090_4 1 MKTFFKTLTSAALALILAACGGQKDSAPAASASPSADNGAKKEIVFGTTVGDGDMVK
zv32_4 1 MKTFFKTLTSAALALILAACGGQKDSAPAASASPSADNGAKKEIVFGTTVGDGDMVK
zv33_4 1 MKTFFKTLTSAALALILAACGGQKDSAPAASASPSADNGAKKEIVFGTTVGDGDMVK
zv02_4 1 MKTFFKTLTSAALALILAACGGQKDSAPAASASPSADNGAKKEIVFGTTVGDGDMVK
zv15_4 1 MKTFFKTLTSAALALILAACGGQKDSAPAASASPSADNGAKKEIVFGTTVGDGDMVK
zv03_4ass 1 MKTFFKTLTSAALALILAACGGQKDSAPAASASPSADNGAKKEIVFGTTVGDGDMVK
zv04_4 1 MKTFFKTLTSAALALILAACGGQKDSAPAASASPSADNGAKKEIVFGTTVGDGDMVK
zv05_4 1 MKTFFKTLTSAALALILAACGGQKDSAPAASASPSADNGAKKEIVFGTTVGDGDMVK
zv09_4 1 MKTFFKTLTSAALALILAACGGQKDSAPAASASPSADNGAKKEIVFGTTVGDGDMVK
zv16_4 1 MKTFFKTLTSAALALILAACGGQKDSAPAASASPSADNGAKKEIVFGTTVGDGDMVK
zv18_4 1 MKTFFKTLTSAALALILAACGGQKDSAPAASASPSADNGAKKEIVFGTTVGDGDMVK
zv26_4 1 MKTFFKTLTSAALALILAACGGQKDSAPAASASPSADNGAKKEIVFGTTVGDGDMVK
zv28_4 1 MKTFFKTLTSAALALILAACGGQKDSAPAASASPSADNGAKKEIVFGTTVGDGDMVK
zv08_4 1 MKTFFKTLTSAALALILAACGGQKDSAPAASASPSADNGAKKEIVFGTTVGDGDMVK
zv10_4 1 MKTFFKTLTSAALALILAACGGQKDSAPAASASPSADNGAKKEIVFGTTVGDGDMVK
zv25_4 1 MKTFFKTLTSAALALILAACGGQKDSAPAASASPSADNGAKKEIVFGTTVGDGDMVK
zv17_4 1 MKTFFKTLTSAALALILAACGGQKDSAPAASASPSADNGAKKEIVFGTTVGDGDMVK
zv96_4 1 MKTFFKTLTSAALALILAACGGQKDSAPAASASPSADNGAKKEIVFGTTVGDGDMVK
zv06_4 1 MKTFFKTLTSAALALILAACGGQKDSAPAASASPSADNGAKKEIVFGTTVGDGDMVK
zv19_4 1 MKTFFKTLTSAALALILAACGGQKDSAPAASASPSADNGAKKEIVFGTTVGDGDMVK
zv20_4 1 MKTFFKTLTSAALALILAACGGQKDSAPAASASPSADNGAKKEIVFGTTVGDGDMVK
z2491_4 1 MKTFFKTLTSAALALILAACGGQKDSAPAASASPSADNGAKKEIVFGTTVGDGDMVK
zv13_4 1 MKTFFKTLTSAALALILAACGGQKDSAPAASASPSADNGAKKEIVFGTTVGDGDMVK
zv17_4 1 MKTFFKTLTSAALALILAACGGQKDSAPAASASPSADNGAKKEIVFGTTVGDGDMVK
zv01_4 1 MKTFFKTLTSAALALILAACGGQKDSAPAASASPSADNGAKKEIVFGTTVGDGDMVK
zv07_4 1 MKTFFKTLTSAALALILAACGGQKDSAPAASASPSADNGAKKEIVFGTTVGDGDMVK
zv21_4 1 MKTFFKTLTSAALALILAACGGQKDSAPAASASPSADNGAKKEIVFGTTVGDGDMVK
zv11_4 1 MKTFFKTLTSAALALILAACGGQKDSAPAASASPSADNGAKKEIVFGTTVGDGDMVK
zv29_4 1 MKTFFKTLTSAALALILAACGGQKDSAPAASASPSADNGAKKEIVFGTTVGDGDMVK
zv22_4 1 MKTFFKTLTSAALALILAACGGQKDSAPAASASPSADNGAKKEIVFGTTVGDGDMVK
zv12_4ass 1 MKTFFKTLTSAALALILAACGGQKDSAPAASASPSADNGAKKEIVFGTTVGDGDMVK
zv24_4ass 1 MKTFFKTLTSAALALILAACGGQKDSAPAASASPSADNGAKKEIVFGTTVGDGDMVK

```

```

fa1090_4 61 DQIQAELEKKGYTVKLVEFTDYVRNLAALAEGLDINVFQHKPYLDDFKKEHNLDITEVF
zv32_4 61 DQIQAELEKKGYTVKLVEFTDYVRNLAALAEGLDINVFQHKPYLDDFKKEHNLDITEVF
zv33_4 61 DQIQAELEKKGYTVKLVEFTDYVRNLAALAEGLDINVFQHKPYLDDFKKEHNLDITEVF
zv02_4 60 DQIQAELEKKGYTVKLVEFTDYVRNLAALAEGLDINVFQHKPYLDDFKKEHNLDITEVF
zv15_4 60 DQIQAELEKKGYTVKLVEFTDYVRNLAALAEGLDINVFQHKPYLDDFKKEHNLDITEVF
zv03_4ass 60 DQIQAELEKKGYTVKLVEFTDYVRNLAALAEGLDINVFQHKPYLDDFKKEHNLDITEVF
zv04_4 60 DQIQAELEKKGYTVKLVEFTDYVRNLAALAEGLDINVFQHKPYLDDFKKEHNLDITEVF
zv05_4 60 DQIQAELEKKGYTVKLVEFTDYVRNLAALAEGLDINVFQHKPYLDDFKKEHNLDITEVF
zv09_4 60 DQIQAELEKKGYTVKLVEFTDYVRNLAALAEGLDINVFQHKPYLDDFKKEHNLDITEVF
zv16_4 60 DQIQAELEKKGYTVKLVEFTDYVRNLAALAEGLDINVFQHKPYLDDFKKEHNLDITEVF
zv18_4 60 DQIQAELEKKGYTVKLVEFTDYVRNLAALAEGLDINVFQHKPYLDDFKKEHNLDITEVF
zv26_4 60 DQIQAELEKKGYTVKLVEFTDYVRNLAALAEGLDINVFQHKPYLDDFKKEHNLDITEVF
zv28_4 60 DQIQAELEKKGYTVKLVEFTDYVRNLAALAEGLDINVFQHKPYLDDFKKEHNLDITEVF
zv08_4 60 DQIQAELEKKGYTVKLVEFTDYVRNLAALAEGLDINVFQHKPYLDDFKKEHNLDITEVF
zv10_4 60 DQIQAELEKKGYTVKLVEFTDYVRNLAALAEGLDINVFQHKPYLDDFKKEHNLDITEVF
zv25_4 60 DQIQAELEKKGYTVKLVEFTDYVRNLAALAEGLDINVFQHKPYLDDFKKEHNLDITEVF
zv17_4 60 DQIQAELEKKGYTVKLVEFTDYVRNLAALAEGLDINVFQHKPYLDDFKKEHNLDITEVF
zv96_4 60 DQIQAELEKKGYTVKLVEFTDYVRNLAALAEGLDINVFQHKPYLDDFKKEHNLDITEVF
zv06_4 60 DQIQAELEKKGYTVKLVEFTDYVRNLAALAEGLDINVFQHKPYLDDFKKEHNLDITEVF
zv19_4 60 DQIQAELEKKGYTVKLVEFTDYVRNLAALAEGLDINVFQHKPYLDDFKKEHNLDITEVF
zv20_4 60 DQIQAELEKKGYTVKLVEFTDYVRNLAALAEGLDINVFQHKPYLDDFKKEHNLDITEVF
z2491_4 60 DQIQAELEKKGYTVKLVEFTDYVRNLAALAEGLDINVFQHKPYLDDFKKEHNLDITEVF
zv13_4 60 DQIQAELEKKGYTVKLVEFTDYVRNLAALAEGLDINVFQHKPYLDDFKKEHNLDITEVF
zv17_4 60 DQIQAELEKKGYTVKLVEFTDYVRNLAALAEGLDINVFQHKPYLDDFKKEHNLDITEVF
zv01_4 60 DQIQAELEKKGYTVKLVEFTDYVRNLAALAEGLDINVFQHKPYLDDFKKEHNLDITEVF
zv07_4 60 DQIQAELEKKGYTVKLVEFTDYVRNLAALAEGLDINVFQHKPYLDDFKKEHNLDITEVF
zv21_4 60 DQIQAELEKKGYTVKLVEFTDYVRNLAALAEGLDINVFQHKPYLDDFKKEHNLDITEVF
zv11_4 60 DQIQAELEKKGYTVKLVEFTDYVRNLAALAEGLDINVFQHKPYLDDFKKEHNLDITEVF
zv29_4 60 DQIQAELEKKGYTVKLVEFTDYVRNLAALAEGLDINVFQHKPYLDDFKKEHNLDITEVF
zv22_4 60 DQIQAELEKKGYTVKLVEFTDYVRNLAALAEGLDINVFQHKPYLDDFKKEHNLDITEVF
zv12_4ass 60 DQIQAELEKKGYTVKLVEFTDYVRNLAALAEGLDINVFQHKPYLDDFKKEHNLDITEVF
zv24_4ass 60 DQIQAELEKKGYTVKLVEFTDYVRNLAALAEGLDINVFQHKPYLDDFKKEHNLDITEVF

```

```

fa1090_4 121 QVPTAPLGLYPGKLKSLLEEVKDGSTVSAPNDPSNFARLVLMLNLGWIKLKDGINPLTAS
zv32_4 121 QVPTAPLGLYPGKLKSLLEEVKDGSTVSAPNDPSNFARLVLMLNLGWIKLKDGINPLTAS
zv33_4 121 QVPTAPLGLYPGKLKSLLEEVKDGSTVSAPNDPSNFARLVLMLNLGWIKLKDGINPLTAS

```

FIG. 21A

DQIQAELEKKGYTVKLVEFTDYVRNLAALAEGLDINVFQHKPYLDDFKKEHNLDITEVF

fa1090_4	181	KADIAENLNKNIKIVLEAAQLPRSRADVDFAVNGNGYAISSGGMKLTALFOEPSFAYVNM
zv32_4	181	KADIAENLNKNIKIVLEAAQLPRSRADVDFAVNGNGYAISSGGMKLTALFOEPSFAYVNM
zv33_4	181	KADIAENLNKNIKIVLEAAQLPRSRADVDFAVNGNGYAISSGGMKLTALFOEPSFAYVNM
zv02_4	180	KADIAENLNKNIKIVLEAAQLPRSRADVDFAVNGNGYAISSGGMKLTALFOEPSFAYVNM
zv15_4	180	KADIAENLNKNIKIVLEAAQLPRSRADVDFAVNGNGYAISSGGMKLTALFOEPSFAYVNM
zv03_4 _{ass}	180	KADIAENLNKNIKIVLEAAQLPRSRADVDFAVNGNGYAISSGGMKLTALFOEPSFAYVNM
zv04_4	180	KADIAENLNKNIKIVLEAAQLPRSRADVDFAVNGNGYAISSGGMKLTALFOEPSFAYVNM
zv05_4	180	KADIAENLNKNIKIVLEAAQLPRSRADVDFAVNGNGYAISSGGMKLTALFOEPSFAYVNM
zv09_4	180	KADIAENLNKNIKIVLEAAQLPRSRADVDFAVNGNGYAISSGGMKLTALFOEPSFAYVNM
zv16_4	180	KADIAENLNKNIKIVLEAAQLPRSRADVDFAVNGNGYAISSGGMKLTALFOEPSFAYVNM
zv18_4	180	KADIAENLNKNIKIVLEAAQLPRSRADVDFAVNGNGYAISSGGMKLTALFOEPSFAYVNM
zv26_4	180	KADIAENLNKNIKIVLEAAQLPRSRADVDFAVNGNGYAISSGGMKLTALFOEPSFAYVNM
zv28_4	180	KADIAENLNKNIKIVLEAAQLPRSRADVDFAVNGNGYAISSGGMKLTALFOEPSFAYVNM
zv08_4	180	KADIAENLNKNIKIVLEAAQLPRSRADVDFAVNGNGYAISSGGMKLTALFOEPSFAYVNM
zv10_4	180	KADIAENLNKNIKIVLEAAQLPRSRADVDFAVNGNGYAISSGGMKLTALFOEPSFAYVNM
zv25_4	180	KADIAENLNKNIKIVLEAAQLPRSRADVDFAVNGNGYAISSGGMKLTALFOEPSFAYVNM
zv17_4	180	KADIAENLNKNIKIVLEAAQLPRSRADVDFAVNGNGYAISSGGMKLTALFOEPSFAYVNM
zv96_4	180	KADIAENLNKNIKIVLEAAQLPRSRADVDFAVNGNGYAISSGGMKLTALFOEPSFAYVNM
zv06_4	180	KADIAENLNKNIKIVLEAAQLPRSRADVDFAVNGNGYAISSGGMKLTALFOEPSFAYVNM
zv19_4	180	KADIAENLNKNIKIVLEAAQLPRSRADVDFAVNGNGYAISSGGMKLTALFOEPSFAYVNM
zv20_4	180	KADIAENLNKNIKIVLEAAQLPRSRADVDFAVNGNGYAISSGGMKLTALFOEPSFAYVNM
z2491_4	180	KADIAENLNKNIKIVLEAAQLPRSRADVDFAVNGNGYAISSGGMKLTALFOEPSFAYVNM
zv13_4	180	KADIAENLNKNIKIVLEAAQLPRSRADVDFAVNGNGYAISSGGMKLTALFOEPSFAYVNM
zv27_4	180	KADIAENLNKNIKIVLEAAQLPRSRADVDFAVNGNGYAISSGGMKLTALFOEPSFAYVNM
zv01_4	180	KADIAENLNKNIKIVLEAAQLPRSRADVDFAVNGNGYAISSGGMKLTALFOEPSFAYVNM
zv07_4	180	KADIAENLNKNIKIVLEAAQLPRSRADVDFAVNGNGYAISSGGMKLTALFOEPSFAYVNM
zv21_4	180	KADIAENLNKNIKIVLEAAQLPRSRADVDFAVNGNGYAISSGGMKLTALFOEPSFAYVNM
zv11_4	180	KADIAENLNKNIKIVLEAAQLPRSRADVDFAVNGNGYAISSGGMKLTALFOEPSFAYVNM
zv29_4	180	KADIAENLNKNIKIVLEAAQLPRSRADVDFAVNGNGYAISSGGMKLTALFOEPSFAYVNM
zv22_4	180	KADIAENLNKNIKIVLEAAQLPRSRADVDFAVNGNGYAISSGGMKLTALFOEPSFAYVNM
zv12_4 _{ass}	180	KADIAENLNKNIKIVLEAAQLPRSRADVDFAVNGNGYAISSGGMKLTALFOEPSFAYVNM
zv24_4 _{ass}	180	KADIAENLNKNIKIVLEAAQLPRSRADVDFAVNGNGYAISSGGMKLTALFOEPSFAYVNM

FIG. 21B

0-9

DOCKET NO. CHIR-0160
(356.001)

PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In Re Application of:

Enzo Scalato, Vega Masignani, Rino Rappuoli,
Mariagrazia Pizza,, and Guido Grandi

Group Art Unit: not assigned

Examiner: not assigned

For: Neisserial Antigens

DECLARATION AND POWER OF ATTORNEY

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name; and

I believe that I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a



Utility Patent



Design Patent

is sought on the invention, whose title appears above, the specification of which:



is attached hereto.



was filed on _____ as Serial No. _____.



said application having been amended on _____

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose to the U.S. Patent and Trademark Office all information known to be material to the patentability of this application in accordance with 37 CFR § 1.56.

I hereby claim foreign priority benefits under 35 U.S.C. § 119(a-d) of any **foreign**

DOCKET NO. CHIR-0160
(356.001)

- 2 -

PATENT

application(s) for patent or inventor's certificate listed below and have also identified below any foreign application for patent or inventor's certificate having a filing date before that of any application on which priority is claimed:

Priority Claimed (If X'd)	Country	Serial Number	Date Filed
<input checked="" type="checkbox"/>	<u>PCT</u>	<u>PCT/IB98/01665</u>	<u>October 9, 1998</u>
<input type="checkbox"/>	<u> </u>	<u> </u>	<u> </u>

I hereby claim the benefit under 35 U.S.C. § 120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States application in the manner provided by the first paragraph of 35 U.S.C. § 112, I acknowledge the duty to disclose to the U.S. Patent and Trademark Office all information known to be material to patentability as defined in 37 CFR § 1.56 which became available between the filing date of the prior application and the national or PCT international filing date of this application:

Serial Number	Date Filed	Patented/Pending/Abandoned
<u> </u>	<u> </u>	<u> </u>
<u> </u>	<u> </u>	<u> </u>

I hereby claim the benefit under 35 U.S.C. § 119(e) of any United States provisional application(s) listed below:

Serial Number	Date Filed
<u> </u>	<u> </u>
<u> </u>	<u> </u>

I hereby appoint the following persons of the firm of **WOODCOCK WASHBURN KURTZ MACKIEWICZ & NORRIS LLP**, One Liberty Place - 46th Floor, Philadelphia, Pennsylvania 19103 as attorneys and/or agents to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith:

Robert B. Washburn	Registration No. 16,574	Michael P. Straher	Registration No. 38,325
Richard E. Kurtz	Registration No. 19,263	David A. Cherry	Registration No. 35,099
John J. Mackiewicz	Registration No. 19,709	Anthony J. Rossi	Registration No. 24,053
Norman L. Norris	Registration No. 24,196	Michael J. Swope	Registration No. 38,041
Albert W. Preston, Jr.	Registration No. 25,366	Michael J. Bonella	Registration No. 41,628
Dale M. Heist	Registration No. 28,425	Harold H. Fullmer	Registration No. 42,560
Philip S. Johnson	Registration No. 27,200	William R. Richter	Registration No. 43,879
John W. Caldwell	Registration No. 28,937	John E. McGlynn	Registration No. 42,863
Gary H. Levin	Registration No. 28,734	Kimberly R. Hild	Registration No. 39,224
Steven J. Rocci	Registration No. 30,489	Lawrence A. Aaronson	Registration No. 38,369
Dianne B. Elderkin	Registration No. 28,598	Jonathan M. Waldman	Registration No. 40,861
Francis A. Paintin	Registration No. 19,386	Paul K. Legard	Registration No. 38,534
John P. Donohue, Jr.	Registration No. 29,916	Chad Ziegler	Registration No. 44,273
Henrik D. Parker	Registration No. 31,863	David N. Farsiou	Registration No. 44,104
Suzanne E. Miller	Registration No. 32,279	Maureen Gibbons	Registration No. 44,121
Lynn B. Morreale	Registration No. 32,842	Steven H. Meyer	Registration No. 37,189
Mark DeLuca	Registration No. 33,229	John M. Paolino	Registration No. 40,340
Joseph Lucci	Registration No. 33,307	Joseph R. Condo	Registration No. 42,431
Michael P. Dunnam	Registration No. 32,611	Michael K. Jones	Registration No. 41,100
Michael D. Stein	Registration No. 34,734	Frank T. Carroll	Registration No. 42,392
Albert J. Marcellino	Registration No. 34,664	Rena Patel	Registration No. 41,412
David R. Bailey	Registration No. 35,057	Mark J. Rosen	Registration No. 39,822
Doreen Yatro Trujillo	Registration No. 35,719	Gregory L. Hillyer	Registration No. 44,154
Barbara L. Mullin	Registration No. 38,250	Maria M. Kourtakis	Registration No. 41,126
Kevin M. Flannery	Registration No. 35,871		
Lynn A. Malinoski	Registration No. 38,788		
Lori Y. Beardell	Registration No. 34,293		

In addition to the attorneys listed above, the undersigned hereby appoints the attorneys listed below of **CHIRON CORPORATION**, 4560 Horton Street, Emeryville, CA 94608-2916 as attorneys for applicants, with full power of substitution and revocation, to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith:

Alisa A. Harbin	Registration No. 33,895
Robert P. Blackburn	Registration No. 30,447
Joseph H. Guth	Registration No. 31,261
Kenneth M. Goldman	Registration No. 34,174.

Please send all future correspondence to:

Alisa A. Harbin, Esq.
Intellectual Property
Chiron Corporation
4560 Horton Street
Emeryville, CA 94608-2916
Telephone: (510) 923-2708

Please direct all telephone calls to:

Mark J. Rosen
WOODCOCK WASHBURN KURTZ
MACKIEWICZ & NORRIS LLP
One Liberty Place - 46th Floor
Philadelphia PA 19103
Telephone No.: **(215) 568-3100**
Facsimile No.: **(215) 568-3439**

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Name: Enzo Scalato	
Mailing Address: Chiron SpA Via Fiorentina 53100 Siena Italy	Signature Date of Signature: _____
City/State of Actual Residence: Colle Val d'Elsa (SI), Italy	Citizenship: <u>Italy</u>

Name: Masignani Vega	
Mailing Address: Chiron SpA Via Fiorentina 53100 Siena Italy	Signature Date of Signature: _____
City/State of Actual Residence: Siena, Italy	Citizenship: <u>Italy</u>

Name: Rino Rappuoli	<hr/> Signature <hr/> Date of Signature: _____ <hr/> Citizenship: <u>Italy</u> _____
Mailing Address: Chiron SpA Via Fiorentina 53100 Siena Italy	
City/State of Actual Residence: Berardenga (SI), Italy	

Name: Mariagrazia Pizza	<hr/> Signature <hr/> Date of Signature: _____ <hr/> Citizenship: <u>Italy</u> _____
Mailing Address: Chiron SpA Via Fiorentina 53100 Siena Italy	
City/State of Actual Residence: Sienna, Italy	

Name: Guido Grandi	<hr/> Signature <hr/> Date of Signature: _____ <hr/> Citizenship: <u>Italy</u> _____
Mailing Address: Chiron SpA Via Fiorentina 53100 Siena Italy	
City/State of Actual Residence: Segrate (MI), Italy	

00303513-0431000